

THE ORGANISATION AND EVOLUTION OF  
A REPEATED DNA SEQUENCE FAMILY  
IN RELATED ALLIUM SPECIES

by

IAN JEFFREY EVANS B.Sc. (NOTTINGHAM)

Thesis submitted to the University of Nottingham  
for the degree of  
Doctor of Philosophy

October 1983

## ACKNOWLEDGEMENTS

In thanking all those members of the department, past and present, who have provided help, encouragement and humour, I would like to single out Dr. S. Barnes, my mentor during both undergraduate and postgraduate years.

In addition, I am very grateful to my parents for their encouragement and support over the years, to Mrs. Rosemary Reid for the considerable effort involved in the typing of the manuscript and to Miss Fiona Benson, without whom the whole thing would not have been quite so much fun.

Acknowledgement is made to the S.E.R.C. for financial support during the period 1980-1983.

## CONTENTS

Contents	.....	i
Summary	.....	iv
Abbreviations	.....	v
 <u>CHAPTER 1 INTRODUCTION:-</u>		
The DNA-sequence organisation of the eukaryotic genome:		
i) Genome organisation as revealed by re-association studies	.....	1
ii) Genome organisation as revealed by cloning studies	.....	13
<u>CHAPTER 2 MATERIALS AND METHODS</u>	.....	28
<u>CHAPTER 3 THE GENUS ALLIUM</u>		
i) Species classification	.....	41
ii) The chromosomes and DNA of <u>Allium</u> species	.....	46
<u>CHAPTER 4 THE CHARACTERISATION OF B1000 IN THE A. SATIVUM GENOME</u>		
i) Restriction endonuclease digestion of <u>A. sativum</u> DNA	.....	52
ii) The cloning of a repeated DNA sequence from the <u>A. sativum</u> genome	.....	54
iii) B1000 sequence is found in a number of genomic configurations	.....	56
iv) The cloned B1000 sequence is a representative member of the genomic sequence family	.....	58
v) Estimation of the repetition frequency of the B1000 sequence in the <u>A. sativum</u> genome	.....	59

- vi) The copy number and physical organisation of B1000 does not appear to be tissue specific ..... 62
- vii) The length of the B1000-associated repetitive element ..... 64
- viii) B1000 sequences are not abundantly transcribed ..... 69

## CHAPTER 5 STUDIES ON CLONED SEQUENCES FROM A. SATIVUM

- i) The isolation and restriction mapping of clones containing B1000-related sequences ..... 71
- ii) The sequence environment of B1000-related sequences ..... 77

## CHAPTER 6 THE CHARACTERISATION OF SEQUENCES HOMOLOGOUS TO B1000 IN RELATED ALLIUM SPECIES

- i) The repetitive sequence components of related Allium species ..... 86
- ii) The prevalence of B1000 in other Allium species ..... 88
- iii) The organisation of B1000 in other Allium species ..... 90
- iv) An estimate of B1000 sequence divergence between species ..... 94

## CHAPTER 7 A SATELLITE-DNA COMPONENT OF A. VINEALE

- i) Identification, isolation and characterisation of a satellite component in A. vineale ..... 100
- ii) The distribution of the 400 bp satellite in related Allium species ..... 103



iii) The organisation of homologous sequences in related <u>Allium</u> species .....	104
Overview and discussion ... ..	106 et. seq.
<b><u>CHAPTER 8 THE EVOLUTION OF REPETITIVE DNA SEQUENCE COMPONENTS</u></b>	
i) Possible mechanisms for the generation of organisational patterns .....	109
ii) The maintenance of patterns and sequences .....	124
<b><u>CHAPTER 9 FUNCTIONS AND EFFECTS ATTRIBUTABLE TO REPETITIVE SEQUENCE DNA</u></b>	132
i) Attributable function .....	133
ii) Attributable effects .....	142
<b><u>CHAPTER 10 REPETITIVE DNA AND THE EVOLUTION OF SPECIES</u></b>	151
<b><u>CHAPTER 11 CONCLUSIONS AND SPECULATIONS</u></b>	163
References .....	171
Publications .....	194

## SUMMARY

A large proportion of the genomes of species belonging to the genus Allium comprises repetitive sequence DNA, a component implicated as a cause of the large variation in C-values between even closely related species.

The work presented here represents part of the first phase in the characterisation of some of these repetitive sequences in a number of Allium species. One repetitive DNA sequence family, B1000, isolated from the genome of A. sativum, has been characterised with respect to the genomic organisation, reiteration frequency and sequence divergence of its members within A. sativum. Sequences sharing homology with a cloned representative member of the B1000 family have been detected in the genomes of a number of other Allium species; such sequences display quantitative and qualitative modulations in their organisation.

In addition, and by contrast, the distribution and organisation of a satellite DNA family present in a number of Allium species has been investigated; the characteristics of this family differ from those of the B1000 family in many respects.

Data relating to the evolution and maintenance, functions and effects of repetitive-sequence DNA in eukaryotic genomes are reviewed and where possible the data pertaining to Allium are discussed in context with such information from other species.

ABBREVIATIONS

bp	base pairs
Brij 58	polyoxyethylene 20 cetyl ether
BSA	bovine serum albumin
Cot	DNA concentration x time (see page 38)
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
EDTA	Ethylenediaminetetra-acetic acid
HAP	hydroxyapatite
kb	Kilobase
LB	Luria broth
LPI	Long-period interspersions
PB	phosphate buffer (see page 38)
PVP	polyvinylpyrrolidone
RNase	ribonuclease
SDS	sodium dodecyl sulphate
SPI	short-period interspersions
SSC	0.15M NaCl + 0.015M Na citrate
TCA	Trichloroacetic acid
Tris	Tris (hydroxymethyl) aminomethane
UV	ultraviolet light

## **CHAPTER I**

### **INTRODUCTION**

## INTRODUCTION

### The DNA-sequence organisation of the eukaryotic Genome

The haploid DNA content (C-value) of eukaryotic genomes ranges from approximately 0.005 pg to 200 pg (Bennett and Smith, 1976), which not only seems to be surplus to coding requirements in most individuals (see for example Judd et al. 1972,; Ohno 1971) but also appears not to correlate with organisational or evolutionary complexity. Although there does appear to be a correlation between the minimum C-value observed in many taxonomic classes and 'complexity' (Sparrow and Nauman, 1976), C-values are nevertheless often very different between even closely related species, a phenomenon referred to as the 'C-value paradox'.

In the hope that a physical characterisation of the complex population of sequences making up eukaryotic genomes will enable the resolution of this paradox, the organisation of DNA sequences in the genomes of many diverse species has been investigated. As a consequence, much information has been collected in the last decade relating to the organisation of eukaryotic genomic DNA sequences at a molecular level.

Models implying an increase in gene number or gene amplification as the major causative factor in DNA content variation (see for example Callan, 1967) have been largely

discredited. Estimates of gene copy number suggested that they remained constant at a low copy number despite variation in genome size. Although there are instances of multiple copy genes (e.g. rDNA and histone genes), most coding sequences appear to be of single copy in the eukaryotic genome (Lewin, 1980).

Interestingly even the amount of single-copy DNA in a genome does not seem to correlate with evolutionary complexity; presumably not all such DNA is coding. Consequently neither the C-value nor the amount of single copy DNA in a genome can be taken as any guide to the number of genes contained within.

The nature of those sequences responsible for the variation in C-values was first suggested by early DNA:DNA solution reassociation experiments in which Britten and Kohne (1968) introduced the concept of Cot curves. A comprehensive review of the methodology and interpretation of such experiments is provided by Britten et al. (1974).

Genomes are composed of highly repetitive sequences, such as satellite DNA, which are found in the non-coding regions of the genome. These sequences are often found in the centromeric and pericentromeric regions of chromosomes. The repetitive nature of these sequences is often reflected in their high C-values.

From Cot analysis, a number of parameters can be derived, such as the C<sub>0</sub>t value, which is a measure of the complexity of the DNA. The C<sub>0</sub>t value is a function of the initial concentration of DNA and the time taken for half of the DNA to reassociate. The C<sub>0</sub>t value is a measure of the complexity of the DNA, and is often used to estimate the size of the genome.

The large volume of data generated using Cot analysis has resulted in the identification of a large number of repetitive sequences in the genome. These sequences are often found in the non-coding regions of the genome, and are often found in the centromeric and pericentromeric regions of chromosomes.

### Genome organisation as revealed by reassociation studies

The majority of eukaryotic DNAs reassociate over a much wider range of Cot values than do prokaryotic standards. This suggests that unlike the almost entirely single copy bacterial genome, much eukaryotic DNA consists of more than one component and includes sequences repeated many times.

Such experiments have formed the basis upon which a great many genomes from many diverse species have been divided into components, each defined by the repetition frequency of its constituent sequences. Such components are rarely discrete and some care is needed in interpreting Cot data.

As a generalisation, eukaryotic genomes typically, but by no means universally, display three components. A 'slow' component renaturing at Cot values greater than about  $10^2$ , comprises single-copy or few-copy sequences. A 'fast' component reassociating before about Cot  $10^{-2}$  comprises highly repetitious sequences such as satellite and fold-back DNA while an intermediate component, renaturing between these two, comprises the rather nebulous 'moderately repetitive' DNA sequences.

From Cot analysis a number of parameters can be estimated, among them the proportion and reiteration frequency of repetitive DNA sequences within a genome. The large volume of data gathered using such techniques has resulted in the characterisation, albeit at a fairly

crude level, of a large number of both animal and plant genomes. Not all studies are directly comparable but most show repetitive DNA to be a very common feature of eukaryotic genomes, in a wide range of species. While Cot analysis provides estimates of the frequency of repetitive sequences, alone it provides no information regarding their length, or arrangement with other sequences.

More detailed information is provided by 'R versus L curves' or 'interspersion curves', first introduced by Davidson et al. (1973) and Graham et al. (1974). The rationale, methodology and interpretation of such experiments is discussed in detail by Davidson et al. (1973), Graham et al. (1974), Rimpau et al. (1978) and Moyziz et al. (1981). Such experiments are particularly informative when used in conjunction with other techniques such as thermal denaturation, hyperchromic shift, S1 nuclease resistance and EM studies, as demonstrated by work detailed later.

Although as in most reassociation experiments there are many complicating factors affecting interpretation, this type of approach not only provides information regarding the nature of the repetitive elements, i.e. their length and the proportion of the genome they make up, but also of the length of the non-repetitive 'spacer' sequences often associated with them and the proportion of the genome consisting of such an 'interspersed' pattern.



The earliest work using these methods revealed that the genomes of Xenopus (Davidson et al. 1973) and sea-urchin (Graham et al., 1974) appeared to be composed, to a large degree, of short repeated sequences (300-400bp) interspersed with single copy sequences of about 1000 bp. Many other genomes were subsequently shown to possess such a short-period interspersion (SPI) pattern.

That such a pattern is however not universal among eukaryotes was first demonstrated in Drosophila (Manning et al. 1975; Crain et al. 1976) in which long (> 5 kb) repetitive sequences appeared to be interspersed with even larger (> 13 kb) single copy sequences in an arrangement known as long-period interspersion (LPI).

Among the reptiles, amphibians and mammals the SPI pattern appeared to be predominant (Epplen et al., 1979; Schmidtke et al., 1979; Schmid and Deininger, 1975) while in birds LPI was more in evidence (Epplen et al., 1978; Eden and Hendrick, 1978). A number of insect genomes appeared to display an LPI component (Crain et al., 1976; French and Manning 1980; Wells et al., 1976) while others were found to be predominantly SPI (Efstratiadis et al. 1976, French and Manning 1980). At a lower level of evolutionary complexity, Fungi and protista seem to be less amenable to classification, some species possessing little or no interspersion (Timberlake, 1978; Krumlauf and Marzluf, 1979),

others possessing characteristics of SPI (Firtel and Kindle, 1975) and others of LPI (Hudspeth et al., 1977).

Although detailed studies are sparse and many taxa have not been investigated, it is clear that there is little consistency within taxa; interspersed patterns do not appear to be maintained even to the extent that related organisms retain similar patterns. This might suggest that evolutionary transition from one pattern to the other is very easily achieved and has occurred on many occasions.

Despite being the basis upon which a large number of eukaryotic genomes have been characterised in the last decade, the distinct nature of SPI and LPI patterns has probably been exaggerated. The classification of genomes as either one or the other appears with hindsight to have been somewhat arbitrary. As Bouchard (1982) states, "The early discovery of two apparently distinct modes of moderately repetitive sequence organisation has had a substantial influence on the attitude of subsequent workers analyzing the organisation of various eukaryotic genomes.". Although in retrospect such a dichotomous system does not seem to be adequate, it was a concept adopted when the relatively crude techniques of DNA reassociation were the most powerful available.

The large volume of data obtained by such methods is reviewed by Bouchard (1982), Singer (1982), Lewin (1980)

and Thompson and Murray (1981), the latter being with reference to plant genomes.

In plants, more so perhaps than in animals, the categorisation of genomes into either SPI or LPI patterns is often meaningless. As Flavell et al. (1974) point out, plant genomes not only contain moderately repetitive sequences but are frequently dominated by them. Consequently, patterns of sequence interspersion unique to very large genomes might be expected. For example, repetitive sequences must be interspersed with each other.

Walbot and Goldberg (1979) considered it reasonable to classify all plant genomes which had been studied at that time as SPI pattern, at least to the extent that single-copy segments of typical length (1-2 kb) are flanked by repetitive DNA sequences. There is clearly however considerable variation in the length of the repetitive segments and the nature of their interspersion with one another. Both long and short repetitive DNA sequences are found in plant genomes, often with a continuous length distribution between. Not all the studies, which have been performed in a wide range of monocotyledonous and dicotyledonous plants, are directly comparable. Some of the more detailed investigations are however informative.

In Soybean (Walbot and Goldberg, 1979; Gurley et al., 1979), approximately 40% of the genome is single-copy sequence, the majority of this being approximately 1 kb

in length and interspersed with repetitive elements of 300-400 bp in a typical SPI organisation. Some longer single-copy DNA is however in an LPI organisation and individual repeats in this species appear to range from 500 bp to several Kilobases in length. The overall organisation can really only be described as showing a general similarity to that in Xenopus.

In pea (Murray et al., 1978), there is again a broad distribution of repetitive element lengths, and a broad range of their reiteration frequencies. The majority of single copy segments appear to be only approximately 300 bp in length, shorter than in the typical SPI organisation.

Fairly extensive investigations, using reassociation techniques, have also been performed on DNA from, for example, Cotton (Walbot and Dure, 1976), Mungbean (Murray et al., 1979), Corn (Hake and Walbot, 1980), Lilium (Bouchard and Stern, 1980), Tobacco (Zimmerman and Goldberg, 1977), Parsley (Kiper and Herzfeld, 1978), Rice (Gupta et al, 1982), Flax (Cullis, 1981), and Pearl Millet (Wimpee and Rawson, 1979). A more extensive list is provided by Flavell (1982).

More useful perhaps than such comparisons between very distant and consequently therefore very different species might be the information obtained from the comparison of genomes of closely related plant species. Such comparisons have been made between some Lathyrus species (Narayan and Rees, 1976; Narayan and Rees, 1977) and perhaps

most comprehensively of all upon four cereal species.

Early work showed approximately 75% of the wheat genome to be repetitive (Flavell and Smith, 1976) most SI-resistant reassociated duplex molecules being 500-600 bp in length. Interspersion of short repetitive sequences with each other was suggested by hyperpolymer formation kinetics (Flavell and Smith, 1977), although single copy sequences of approximately 1000 bp were also found to be interspersed with such short repetitive sequences.

The genome of rye, which is 50% larger, appears to contain, in addition to sequences of 500-600 bp, very long repetitive elements. Most of the additional DNA in rye consists of sequences also present in wheat, suggesting that amplification of sequences may be important in evolution. Single copy sequences, mostly <1500 bp, are interspersed in rye with repetitive sequences of weight average 680 bp (Smith and Flavell, 1977), although since repeated sequence DNA constitutes such a large proportion of the rye genome (about 70%), a typical SPI pattern is impossible; long regions of repetitive sequences must exist in the genome.

The organisation of those sequences which are sufficiently closely related to form interspecies heteroduplexes was investigated by the method of interspersion curve construction; different length tracer

fragments of wheat DNA were reassociated with driver DNA from wheat and oats (Smith et al., 1976) and the homo- and hetero-duplexes further characterised by S1 nuclease digestion and elution by phosphate or temperature increases from hydroxyapatite. 16-22% of the wheat genome consists of repetitive sequences that are related, albeit somewhat distantly, to repetitive sequences in oats under the standard stringencies employed. These common sequences appear to be distributed throughout 67% of the wheat genome at intervals of >3500 bp. Sequences common to both species are not clustered in the wheat genome and must therefore be interspersed with sequences that are not found in oats.

These studies were extended by reassociating labelled tracer DNA, of different lengths, from each of the four cereal species; rye, oats, wheat and barley with driver DNA of each in all combinations (Flavell et al., 1977). The construction of such interspecies interspersion curves permitted the degree of homology between each of the DNAs to be estimated and a phylogenetic scheme of cereal DNA evolution to be constructed. Seven groups of repetitive DNA sequences were identified on this basis, each of which was present to a differing degree (or indeed absent) in each species. The thermal stability of heteroduplexes suggested that .. "repeated sequences of a family are more closely related within than between species .." which was interpreted as suggesting that a

series of amplification events had taken place during evolution.

The nature and organisation of the seven groups of repetitive sequence was further characterised, each group being defined, distinguished and isolated by its presence or absence in the four genomes. The organisation of each group of sequence was investigated by reassociation using different length tracer fragments of wheat and rye DNA against driver DNA of each group. (Rimpau et al., 1978). The sequence organisation of wheat and rye DNA was described in terms of different units, comprising different groups of sequences, and crude organisational maps constructed. Over 80% of non-repetitive DNA in wheat and rye was found to reside close to a repetitive sequence in a pattern resembling that of SPI. In units lacking single-copy DNA however, short repetitive sequences belonging to different groups appeared to be highly interspersed with each other, suggesting amplification and dispersal or re-amplification of already interspersed repeats, concepts which will be discussed later. Some of the data are presented graphically in figure 1.

Similar studies, leading to similar conclusions, have been performed using barley and oat DNA (Rimpau et al., 1980).

Such studies however appear to take the techniques of DNA:DNA reassociation to their limit. The organisational maps of cereal genomes generated in this way are,

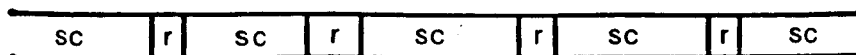
## FIGURE 1

Generalised common units of sequence organisation in plant genomes as revealed by reassociation studies.

- A. Interspersion of single-copy sequences with short repetitive elements.
- B. Interspersion of members of different repetitive families of heterogeneous length.
- C. Tandem arrays of homologous members of one family of repetitive sequences.
- D. Long regions of uninterrupted single-copy sequences.

For details of specific examples in cereal genomes see Rimpau et al. (1978).





A



B



C



D

sc = single copy DNA  
r = repetitive DNA

in the author's words, "simplified and schematic .." and display .. "inadequacies with respect to fine detail.."

Moyzis et al. (1981) extended and modified the biochemical and biophysical methods of Davidson et al. (1973) and Graham et al. (1974), applying them to an investigation into repetitive sequence interspersion in Syrian hamster DNA. In so doing the authors suggest four modifications to the techniques routinely employed, without which they suggest that interspersion data can be incorrectly interpreted. They further suggest that in previous studies the amount of SPI may have been overestimated and that single copy regions in Syrian hamster, rat and human DNA are much longer than the generally accepted 1 kb.

Perhaps the most important observation among the confusion of interpretations is that the vast majority of eukaryotic genomes contain relatively short moderately repetitive sequences that are interspersed with unrelated sequences, whether they be single-copy or other repeats. Although individual variation may be common, the phenomenon of generation and distribution of families of repeated sequences seems to be a common one. A knowledge of the physical organisation of DNA in a eukaryotic genome is essential if the mechanism responsible for such a phenomenon is to be elucidated. This in turn may suggest a function for such sequences, if one exists.

### Genome organisation as revealed by cloning studies

The problem with reassociation related work is that an average picture is presented which may obscure finer details of genome organisation. In the first decade of such work it was not possible easily to isolate a pure sequence-family from the complex population of genomic sequences. What made such isolation possible was the advent of molecular cloning. The use of pure cloned repetitive sequence DNAs as probes in conjunction with such techniques as restriction mapping and DNA sequencing has led to the characterisation of a large number of repetitive sequences at a very fine level in a wide range of species. Reassuringly, in a number of instances, detailed studies at a molecular level have confirmed properties of elements that had been predicted from earlier reassociation studies. One such is the nature of Alu sequences.

Alu sequences were first reported by Houck et al. (1979) as 300 bp duplex molecules resistant to S1 nuclease digestion after reassociation of sheared denatured human DNA. Reassociation experiments demonstrated that they were interspersed in the human genome with non-related sequences, many adjacent to non-repetitive DNA, and that they had a reiteration frequency of as many as  $3 \times 10^5$ , representing 3% of the genome. Furthermore, some members of this family

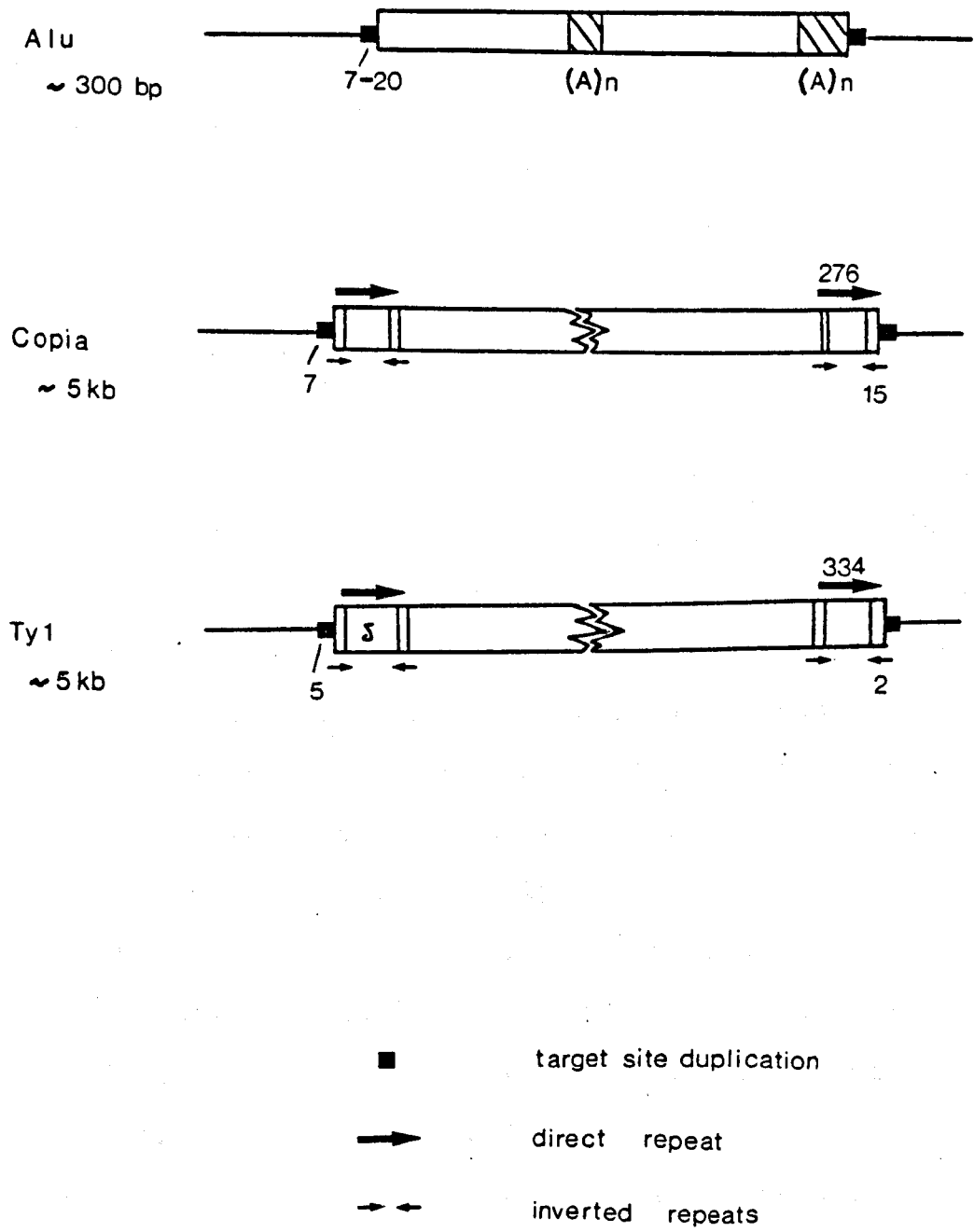
appeared to be transcribed, being found on HnRNA and cytoplasmic poly A<sup>+</sup> RNA (Jelinek et al., 1980).

Following the cloning of members of the Alu family, detailed analysis showed that most elements were indeed 300 bp in length and that they were interspersed throughout most of the human genome, as evidenced by the fact that they are present in 90% of the clones in a genomic  $\lambda$  library (Tashima et al., 1981). Alu elements have been found in the region of known genes and within gene clusters (Bell et al., 1980; Duncan et al., 1979; Fritsch et al., 1980). A consensus sequence has been constructed (Deininger et al., 1981) to which the majority of family members show at least 80% sequence homology. An internal structure is also evident, all elements being a head-to-tail dimer; each half being terminated by an AT rich sequence and the whole being flanked by short direct repeats (see figure 2 ).

Many Alu elements act as in vitro and in vivo templates for transcription by RNA polymerase III (Duncan et al., 1979; Elder et al., 1981). This, in addition to the fact that such elements are often found within such usually well conserved and homogeneous sequences as satellite arrays (Grimaldi and Singer, 1982), has been interpreted as suggesting that Alu elements are capable of transposition.

FIGURE 2

The general structure of the eukaryotic repetitive elements Alu, copia and Ty1.



Supporting this theory is the fact that genomic Alu elements are often found surrounded by short direct repeats, duplications of a genomic sequence, which is typical of bacterial transposons and, as is becoming increasingly apparant, also typical of eukaryotic transposons. A possible mechanism for transposition is the reverse transcription and reinsertion of transcribed Alu elements (Jagadeeswaran et al., 1981; Sharp, 1983), the presence of Alu elements on polydispersed circular DNA from monkey cell cultures perhaps being evidence of an intermediate in this process (Krolewski et al., 1982).

Alu-like sequences are not confined to the human genome. They have also been reported in other primate genomes such as the bonnet monkey (Houck and Schmid, 1981), African Green monkey (Grimaldi et al., 1981) and the Galago (Houck and Schmid, 1981), where they are also interspersed and 300 bp in length.

Related sequences of about 130 bp have been reported in rodent genomes. The B1 and B2 sequences in the mouse are strikingly similar to primate Alu sequences in copy number, interspersion pattern, orientation in the genome and transcription (Kramero, et al., 1979; Georgiev et al., 1981; Krayev et al., 1980). The presence of Alu-like sequences in the genome of a plant species, Maize, has been demonstrated

(Blin et al., 1983). Several other short interspersed repetitive sequences that are unrelated to Alu have been described, among them Ret (875 bp) in African Green Monkey (Lewin, 1982) and A, B and C in mouse (Haigwood et al., 1981).

It seems reasonable to assume that it is the presence of such short elements, which are usually in high copy number and widely interspersed in a genome, that is responsible for the reassociation of DNA fragments in a manner suggesting SPI in some of the eukaryotic genomes alluded to earlier.

That many sequences in the eukaryotic genome appeared to be transposable was perhaps surprising, although earlier studies had suggested that such a phenomenon might be responsible for mutations of high reversion frequency such as those in Maize (McClintock, 1956). Transposition of sequences was also invoked, without any direct evidence, for the interspersion patterns revealed by reassociation work. It was only with detailed genomic mapping that elements capable of transposition were characterised. One of the earliest and better characterised were the copia-like sequences in Drosophila.

Copia and the copia-like sequences such as 412 and 297 (Finnegan et al., 1982), mdg 1 and mdg 3 (Ilyin et al., 1980) share a number of properties. All are much longer than Alu-like elements, being

approximately 5-7 kb, exist in 30-50 copies per genome, are interspersed and are flanked by long (200-500 bp) direct repeats; see figure 2.

Members of each family code for abundant poly A<sup>+</sup> mRNA. The resemblance to transposable elements of E. coli (Calos and Miller, 1980) and to the proviruses of vertebrate retroviruses (Levis et al., 1980) has been noted.

The copy number and chromosomal location of many of these elements appears to vary between different Drosophila strains (Strobel et al., 1979), there being a short duplication of genomic DNA sequence associated with each element. Furthermore, a number of revertable mutations have been shown to be due to the insertion and excision of copia-like transposable elements (Gehring and Paro, 1980; Snyder et al., 1982). The evidence strongly suggests that copia-like repetitive sequences are transposable and although the mechanism is not clear, the existence of copia-like RNA sequences in retrovirus-like particles in cultured D. melanogaster cells (Shiba and Saigo, 1983) is suggestive.

An unusual family of transposable elements in Drosophila has been described by Truett et al. (1981). The foldback (FB) family members are heterogeneous in both the length of the element and the length of the terminal inverted repeats that characterise them.



There are significant differences in chromosomal locations between strains, there usually being about 30 copies per genome. An FB element responsible for the  $W^C$  mutation at the white locus in Drosophila is capable of precise excision at high frequencies. (Collins and Rubin, 1983).

Transposable elements have also been described in other eukaryotic genomes. Tyl elements of Saccharomyces cerevisiae have properties very similar to those of copia (Cameron et al., 1979) and have been shown to be responsible for an unstable mutation (Roeder et al., 1980). The Caenorhabditis elegans genome contains a 1.7 kb transposable element, Tc1 (Emmons et al., 1983) while in plants a 402 bp element flanked by direct repeats has been detected in the genome of Maize (Marx, 1983).

Many other classes of repetitive DNA sequence elements in eukaryotes have been described. Evidence for their being transposable is however suggestive rather than compelling. The more extensively investigated of these are probably the Kpn and Hind III 1.9 kb families in primates, the EcoRI/MIF-1 family in rodents and a number of families in sea urchin.

Six distinct families of Kpn I sequences of the human genome have been described (Shafit-Zagardo, 1983), members of which are polymorphic in length and sequence, occurring approximately  $10^4$  times per genome and which are found interspersed with other sequences including genes and satellite arrays. Some members

appear to be transcribed (Lerman et al., 1983). Members of the 1.9 kb Hind III family of repeats in human DNA appear to be interspersed as part of a larger (> 5.5 kb) element repeated approximately 3000-4000 times (Manuelidis and Biro, 1982). They appear to be related to the above-mentioned Kpn I sequences.

MIF-1 family members, of which there are approximately 20,000 interspersed throughout the genomes of several rodent species (Brown and Dover, 1981), are at least 3000 bp in length. They appear to be the same sequence as the 1.3 kb Eco RI family (Heller and Arnheim, 1980; Cheng and Schildkraut, 1980; Meunier-Rotival et al., 1982), members of which have been investigated in some detail and appear to be somewhat heterogeneous with respect to sequence and length.

Members of specific repetitive families in the genome of the sea urchin Stroglyocentrotus purpuratus have been characterised with respect to their organisation (Anderson et al., 1981; Moore et al., 1981; Scheller et al., 1981; Posakony et al., 1981). The results are complex but both long and short repetitive elements are described. Interspersion of such elements with both single copy DNA and other repetitive elements is common.

An arrangement of repeated DNA sequences previously undetected in Drosophila was revealed by cross-hybridisation of fragments of ten clones (Wensink et al., 1979). Repetitive elements no greater than 1 kb in length were found to reside in large clusters with unrelated elements of similar size. Different clusters share the same short elements but such elements are in different permutations in each cluster. This 'clustered and scrambled' arrangement of moderately repetitive elements suggests that they may be transposable, as indeed might be entire clusters.

A clustered organisation in which repetitive-sequence elements are scrambled at different sites in the genome has also been detected in the chicken genome (Musti et al., 1981).

Clearly a large proportion of many eukaryotic genomes is composed of moderately repetitive interspersed sequences such as those described above; sequences which are often largely responsible for the differences in DNA content between related species.

In a number of species however a large proportion of the genome comprises highly repetitive tandemly arranged 'satellite-DNA' sequences; 50% for example in Dipodomys (Hatch et al., 1976). The organisation

of satellite DNA in eukaryote genomes is reviewed by John and Miklos (1979), Lewin (1980) and Singer (1982).

The presence of such sequences was originally deduced from the fact that they form 'satellite' bands when subjected to isopycnic centrifugation in CsCl or  $\text{CuSO}_4$  density gradients, as a result of their average base composition deviating significantly from that of total DNA. A number of satellites are 'cryptic', only being revealed in the presence of heavy metal ions antibiotics or dyes in such gradients.

Early reassociation studies established that satellite DNA is generally highly repetitious and in the form of a tandem repeat of a unit sequence. The basic unit may be as small as 2bp (Sueoka N and Cheng T.Y., 1962 ) or be several hundreds of base pairs in length. The term satellite DNA is now applied to DNA sequences of this organisation, rather than to just those sequences isolatable as satellite bands on isopycnic gradients. On the other hand although tandemly repetitive genes, for example r RNA genes, can form satellite bands in such gradients they are usually not regarded as satellite DNA. For the purposes of this discussion such sequences will be referred to as 'repetitive genes'.

In a simple situation, partial restriction digestion of satellite DNA liberates a characteristic

ladder, fragments being integral multiples of the basic unit. The situation is, however, often more complex, restriction digestion often revealing the presence of short and long range periodicity. In addition the existence of more than one type of restriction pattern reveals the presence of discrete segments of variant sequences in many genomes (for example Hörz and Zachau, 1977). In such a way restriction digestion has revealed much regarding the evolutionary history of satellite DNA.

In situ hybridisation studies have demonstrated that satellite DNA sequences are generally associated with heterochromatin, a fact that has provoked much speculation regarding possible functions of such sequences. Ideas relating to the evolutionary history and possible function of satellite DNA will be discussed later.

In evolutionary terms satellite DNA is rather variable. Some species possess characteristic satellite DNA sequences which will not cross reassociate with those sequences of related organisms (for example in some Drosophila species). Indeed some organisms possess a number of satellite families, which may or may not be related to one another, while others may have one, or indeed may lack such a component. In instances where related species share a

satellite DNA family the quantity may be very different between closely related genomes. An example is in Dipodomys species (Hatch et al., 1976).

In contrast to their great variability between related species (either qualitative or quantitative), satellite families show relative uniformity within a species. It has long been assumed that the same relatively homogeneous arrays were present on all chromosomes of the complement in many species. The use of cloned material however has enabled the sequencing of a number of satellite units, revealing a considerable amount of sequence divergence. Whether this is a result of heterogeneity within an array present on all chromosomes or whether it is due to divergence between arrays on different chromosomes, arrays which are homologous within a chromosome, remains to be established. The evolutionary implications of within species variation of satellite DNA arrays will be considered later.

The vast majority of satellite DNA shows no detectable level of transcription. The exceptions (Varley et al., 1980) might be a result of readthrough from upstream promoters.

Relatively few investigations have been made relating to the characterisation in detail of the many repetitive sequences present in plant genomes. The situation however seems to be more complex than

in animal genomes.

Work using cloned repetitive sequences has been mainly centred upon those tandemly arranged highly repeated DNA sequences present in heterochromatin. The organisation of both simple and compound tandem repeats in the genomes of species of the Gramineae has been determined (Bedbrook et al., 1980 a, b; Jones and Flavell, 1982). The localisation, copy number and structure of four families of tandemly arranged simple sequence DNA was investigated in a number of Secale species (Jones and Flavell, 1982). Each species was unique in the amount of each family and its distribution within the chromosome complement, although there was usually an association with heterochromatin. The evidence suggested that each sequence family exists independently of the others, perhaps residing in separate blocks. In contrast, other repetitive sequence families appeared to be conserved between species as evidenced by fluorescence bands on ethidium bromide stained gels.

Reamplification of a portion of a tandem array after the insertion of a short sequence within the array appears to have led to the telomeric-specific compound repeats of rye (Bedbrook et al., 1980 a, b).

Clearly however many families of repeated DNA sequences are organised in a way more complex than a localised tandem array. Many appear to be

interspersed with other non-related repeating units or with single-copy DNA and to be distributed along all chromosomes of the complement. A randomly isolated wheat DNA fragment was found to consist entirely of repeated sequence DNA representing members of several families (Flavell et al., 1981). In situ hybridisation to metaphase wheat chromosomes showed the sequence to be distributed over most if not all the chromosomes.

Results using two such wheat clones as probes against restricted total DNA, which had been fractionated and bound to nitrocellulose filter, showed that members of the family represented by the clones vary considerably in copy number and in organisation between related species (Flavell, 1982). Such a conclusion is drawn from the variation in intensity and in the pattern of hybridisation. The organisation of the repeats responsible for these data is presumably complex, reflecting their evolutionary history. Sequence translocation, amplification and possibly deletion appear to be responsible for the differences in the majority of DNA sequence families between related species.

A system for classifying moderately repetitive interspersed sequences has been suggested by Singer (1982). Although useful in categorising those sequences that have so far been described in eukaryotes,



it may suffer from the same inadequacies as does the partitioning of genomic interspersion patterns into SPI and LPI; In other words there is a danger of subjecting what is effectively a rather heterogeneous situation to a dichotomous classification system.

Nevertheless the concept of 'short interspersed repeated sequences' (SINES) and 'long interspersed repeated sequences' (LINES) has been adopted by a number of authors. SINES are usually considered to be shorter than 500 bp and to have a copy number of up to  $10^6$  per genome. LINES on the other hand are several Kilobases in length and are generally of lower copy number ( $10^4$  or lower). Some of the repetitive elements described above can be fitted into this classification; Alu-like sequences being considered SINES; Kpn, Tyl, MIF-1 and all copia-like sequences as LINES. The 'clustered and scrambled' repeats of Drosophila and the complex repeats of cereal genomes do not fall easily into either category.

Unfortunately even a detailed knowledge of a particular sequence does not necessarily lead immediately to the elucidation of its function or evolutionary behaviour. A prime example is that of satellite DNA. As John and Miklos (1979) conclude in their review of the field ... "although fashionable dogma dictates that analyses of structure and change in structure will necessarily resolve the questions

of function, it is obvious that such oblique approaches do not for the most part even impinge upon the problem ..."

Nevertheless, in the absence of any direct methods of elucidating function and evolutionary behaviour, detailed comparison of the organisation of specific repetitive sequence families between related species remains perhaps the most powerful approach.

It was with this in mind that a study of a class of repetitive DNA sequence elements found in the genome of A. sativum was undertaken; elements which are indeed common to several other Allium genomes.

## CHAPTER 2

### MATERIALS AND METHODS

## MATERIALS AND METHODS

### Materials

Allium material was obtained commercially except for A. ursinum (Bulwell, Notts.), A. babingtonii (Tresco Cornwall), A. triquetrum (St. Marys, Scilly Isles), A. vineale (Portland Bill, Dorset).

### Isolation of plant DNA

Allium tissue (bulb, unless otherwise stated) was ground in a minimum volume of extraction buffer (0.5M Sucrose, 0.05M Tris, 0.1M EDTA, p.H 7.23), strained through gauze or miracloth (Calbiochem) and centrifuged for 15 minutes at 3,000 r.p.m. (4°C). The resulting pellet was resuspended in extraction buffer containing 3.3% Triton-X and centrifuged for 15 minutes at 5,000 r.p.m. This step was repeated several times and the pellet resuspended in 10 volumes of 0.1M Tris, 0.1M EDTA pH 8.0 to which SDS was added to 0.5% and protease (preincubated at 37°C for ½ hr.) to 500 µg/ml.

After one hour incubation at 37°C, 1/5 volume of 6M sodium perchlorate was added to the lysed material and the mixture shaken with an equal volume of chloroform:amylalcohol (23:1). This mixture was centrifuged for five minutes at 5,000 r.p.m., the upper

layer being retained, and the operation repeated until no more material appeared at the interface. An equal volume of cold absolute ethanol was layered on and the DNA spooled into a suitable buffer (e.g. 10mM Tris, 10 mM EDTA, pH 8.4).

#### Purification of plant DNA

DNA extracted by the above method was subjected to two rounds of centrifugation on neutral CsCl density gradients in the presence of ethidium bromide. DNA was centrifuged in gradients comprising 4.275g CsCl, 60  $\mu$ l Ethidium bromide (10 mg/ml) and 4.6 ml DNA in buffer (e.g. 50 mM Tris, 10 mM EDTA, 0.1% sarcosyl) for 16-20 hours at 40,000 r.p.m., 20°C. in a vertical rotor (VTi65) on a Beckman L5 65B ultracentrifuge.

Mainband DNA was visualised in UV light, fractionated by side puncture, shaken ten times with isopropanol to remove ethidium bromide and dialysed overnight at 4°C against a dilute buffer such as 10 mM Tris - HCl pH 8.0.

Purified plant DNA was stored at -20°C.

#### Isolation and purification of plasmid DNA

Large scale isolation of plasmid DNA was performed essentially as described by Clewell and Helinski (1969). 30 mls (or multiples thereof) of Luria Broth was inoculated with 1 ml of an overnight culture and incubated with shaking to OD = 0.6.

Chloramphenicol was then added to 200 µg/ml and incubation continued for 18 hours. Cells were harvested and resuspended in 1ml cold 25% sucrose in 0.05M Tris pH8.0. 1 mg lysozyme, in 0.25M Tris pH 8.0, was added and the cells kept on ice for five minutes. 0.4ml. of 0.25M EDTA pH 8.0 and, after five minutes on ice, 1.6ml detergent mixture, were added to lyse the cells (1% Brij 58, 0.4% sodium deoxycholate, 0.0625M EDTA, 0.05M Tris pH 8.0). After ten minutes gentle agitation the clear viscous mixture was centrifuged for thirty minutes at 20,000 r.p.m., (4°C) to remove cell debris and associated 'chromosomal' DNA. The supernatant was shaken several times with an equal volume of chloroform: amyl alcohol (23:1). Finally an equal volume of cold ethanol was added, the mixture shaken gently, held at -20°C for at least one hour and the DNA precipitate pelleted by centrifugation at 10,000 r.p.m. for ten minutes.

Plasmid DNA was purified by centrifugation on neutral CsCl density gradients in the presence of ethidium bromide as described above, except that the gradients comprised 4.5g CsCl, 60 µl(10mg/ml) ethidium bromide, and 4.5ml DNA in buffer (e.g. 0.1M Tris- HCl, 0.1M EDTA pH 8.4).

Plasmid DNA was also routinely extracted on a smaller scale by the rapid lysis method of Ish-Horowicz and Burke (1981). Cells from 1-4 mls of

overnight culture were harvested and resuspended in 100  $\mu$ l 50mM glucose, 25mM Tris pH 8.0, 10mM EDTA. 200  $\mu$ l 0.2 M NaOH, 1% SDS was added and the mixture incubated on ice for five minutes. Upon addition of 150  $\mu$ l cold 5M NaOAc pH 4.8, a heavy precipitate formed which was pelleted. The supernatant was removed, and the DNA ethanol precipitated and resuspended in 10mM Tris pH 8.0. Contaminating RNA was removed during restriction digests of plasmid DNA by the addition of 2  $\mu$ g/ml heat treated ribonuclease A (Sigma).

#### Restriction digestion

Restriction endonuclease digestion of DNA was performed in an appropriate volume at 37°C for 1½-2 hours with an excess of enzyme usually in a buffer comprising 100mM Tris-HCl (pH 7.2), 5mM MgCl<sub>2</sub>, 50mM NaCl or in BRL core buffer (50mM Tris HCl pH 8.0, 10mM MgCl<sub>2</sub>, 50mM NaCl). Reactions were stopped by the addition of 1/4 volume of 10% ficoll, 0.06% Bromophenol blue, 0.5% SDS and incubation at 65°C for ten minutes.

Restriction endonucleases were obtained from BRL, Enzobiochem Inc., CBL and Biolabs.

#### Gel electrophoresis

DNA fragments were electrophoresed on vertical and horizontal agarose gels in Tris Acetate EDTA

buffer (0.04M Tris, 5mM Na acetate, 1mM EDTA pH 7.9) and on agarose mini-gels (Uniscience) in Tris borate buffer (0.1M Tris, 0.1M Boric acid, 2.5mM EDTA).

The Tris Acetate EDTA buffer was also used for polyacrylamide gel electrophoresis. After staining with Ethidium Bromide (0.5µg/ml), DNA was visualised on a UV transilluminator and photographed on Ilford FP4 film.

#### Isolation of DNA from agarose gels

Stained DNA was cut from the gel in an agarose strip, placed in dialysis tubing containing 1/10 dilution electrophoresis buffer and electroeluted in the same buffer for  $\frac{1}{2}$  hour at 100 v. The buffer was removed from the tubing, the DNA extracted with phenol: chloroform (1:1) and chloroform, and ethanol precipitated after the addition of 1/10 volume 2M Na Acetate.

#### Construction of recombinant plasmids and Transformation

Partial genome libraries were constructed in pBR 322, pAT 153 and pBR 329 by digesting approximately 1µg of genomic DNA and 1µg of plasmid DNA to completion with the appropriate enzyme, mixing the two in a buffer comprising 66mM MgCl<sub>2</sub>, 0.066 mM ATP, 10mM DTT, 66 mM Tris HCl pH 7.6 and ligating for 16-20 hours at 4°C with 0.1 units ligase (BRL).

Plasmid DNA was transformed into calcium-treated



E. coli (HB101) cells. 8ml LB were inoculated with 0.4ml of a fresh overnight culture and incubated at 37°C with agitation for 90-100 minutes. The cells were harvested and resuspended in 4ml cold 10mM CaCl<sub>2</sub>. After reharvesting and resuspension in 0.5ml cold 75mM CaCl<sub>2</sub>, 0.2ml of the cell suspension was added to the DNA in 0.3ml 75mM CaCl<sub>2</sub>. This mixture was held on ice for 45 minutes and heat shocked at 42°C for ten minutes. 0.5ml LB was added and the cells grown at 37°C for 1-2 hours before plating on the appropriate LB plates.

Recombinants were selected on the basis of insertional inactivation of either the ampicillin or tetracycline gene, and stored at -20°C in 35% glycerol.

#### Screening of recombinants

Clones containing a particular DNA sequence were selected by colony filter hybridisation, essentially as described by Grunstein and Hogness (1975) except that Whatman 541 hardened ashless filter paper was used and the proteinase K step omitted.

#### In vitro labelling of DNA by Nick-translation

Nick-translation was carried out essentially according to Rigby et al (1977) and A. Jeffreys (pers.comm.), the 25 µl reaction mixture comprising 0.05M Tris-HCl pH 7.8, 5mM MgCl<sub>2</sub>, 0.01 M

$\beta$ mercaptoethanol, 5  $\mu$ M dATP, 5  $\mu$ M dGTP, 5  $\mu$ M dTTP, 1-2  $\mu$ l ( $\alpha$ - $^{32}$ P) dCTP (10mCi/ml, Amersham) and 0.05-0.1  $\mu$ g DNA. 1  $\mu$ l 8ng/ml DNase and 4 units DNA polymerase I (BRL) were added and the mixture incubated at 15°C for one hour. The reaction was stopped with 25  $\mu$ l 10mM Tris-HCl pH 7.5, 10mM EDTA, 0.5% SDS and DNA extracted with phenol:chloroform (1:1) followed by two ethanol precipitations using Calf Thymus DNA (0.1mg) as carrier DNA. The final precipitate was redissolved in 500  $\mu$ l 10mM Tris HCl pH 7.4 and stored at -20°C.

Specific activities of  $10^7$  -  $10^8$  counts/min per  $\mu$ g of DNA were routinely obtained.

#### In vitro end-labelling of DNA fragments

Restriction fragments of plasmids were ethanol precipitated, resuspended in 10  $\mu$ l H<sub>2</sub>O and were end-labelled in 1 x core buffer (BRL) in the presence of 5  $\mu$ M dATP, 5  $\mu$ M dGTP, 5  $\mu$ M dTTP, 0.5  $\mu$ l ( $\alpha$  - $^{32}$ P)

dCTP (10mCi/ml, Amersham) and 2 units large-fragment (Klenow) polymerase (BRL), essentially by the method of Maniatis et al (1982). The reaction was performed at room temperature (20-24°C) for  $\frac{1}{2}$  hour after which time the mixture was loaded directly onto an agarose mini-gel on which the restriction fragments were fractionated. The required DNA fragment was

excised after staining with ethidium bromide and visualising in UV light, electroeluted, extracted with phenol:chloroform and chloroform, ethanol precipitated and redissolved in 50 $\mu$ l 10mM Tris HCl pH 7.4.

#### DNA transfer and filter hybridisation

DNA was transferred from agarose gels to nitrocellulose filter paper (Schleicher and Schüll BA 85) essentially as described by Southern (1976).

Large (>10kb) fragments of DNA were often partially depurinated by treatment with 0.2M HCl for ten minutes to aid transfer. Following denaturation in 1.5M NaCl, 0.5M NaOH and neutralisation in 3M NaCl, 0.5M Tris HCl pH 7.0, each for 15-30 minutes, DNA was transferred overnight at room temperature. The filter was then washed in 2 x SSC and baked in vacuo at 80°C for 2 hours. Filters were stored at room temperature.

Filters to be probed were prehybridised for several hours at 65°C, in the hybridisation solution (4 x SSC, 0.1% SDS, 5 x 'Denhardtts solution' i.e. 0.1% PVP, 0.1% ficoll, 0.1% BSA) in sealed plastic bags. Heat denatured probe was added directly to this solution (usually in about 100 $\mu$ l) and hybridisation performed for 15-18 hours at 65°C. In some experiments, to increase the rate of reassociation, Dextran

sulphate (high M.W, Sigma) was added to the prehybridisation solution to a final concentration of 10% (Wahl et al, 1979).

Following hybridisation, filters were washed in 4 x SSC, 0.1% SDS at 65°C for at least ½ hour and also in 2 x SSC at room temperature before being blotted dry and subjected to Autoradiography.

Some nitrocellulose filters were re-probed several times. Those filters on which probe had decayed to below detection level were re-probed in the way described above for freshly baked filters. Otherwise probe was removed from a filter by incubation at 65°C in H<sub>2</sub>O for several hours, with several changes, after which time hybridisation was performed as described above.

#### Autoradiography

Filters were exposed to pre-flashed Kodak X-omatS film, usually with an intensifying screen (Fuji Mach II) at -80°C (Laskey, 1980) for periods ranging from 2 hours to several weeks. Film was developed in Kodak DX-80 developer.

#### Spot-blot hybridisation

Samples of DNA were quantitatively assessed before transfer to nitrocellulose filter and hybridisation essentially by the method of Cunningham (1983). 5µl samples of DNA solutions were spotted

onto ethidium bromide plates (1% agarose containing 10<sub>μ</sub>g/ml Ethidium bromide poured in petri dishes) and allowed to dry. DNA was visualised by UV and quantified against a  $\lambda$ -DNA series -dilution standard. A thin layer of 1% agarose was poured over the gel and allowed to set. The DNA was partially depurinated, denatured and neutralised as described above ('DNA transfer and filter hybridisation') and transferred to nitrocellulose paper in a simplified Southern blot apparatus for at least 2 hours. The filter was baked in vacuo for two hours and probed in the usual fashion.

#### Rapid small-scale extraction of plant RNA

Total RNA was extracted from root tissue of a week-old individual plant by an adaptation of the method of Bartels and Thompson (1983). The tissue was finely chopped and ground in a pestle and mortar in a minimum volume of 0.1M NaCl, 0.01M Tris, 0.01M EDTA, 2% SDS. Preincubated protease was added to 2mg/ml and the mixture incubated at 37°C for 15 minutes. Cell debris was pelleted by centrifugation and the supernatant extracted twice with phenol: chloroform (1:1) and with chloroform. Nucleic acid was ethanol precipitated and redissolved in 10mM Tris pH 8.0. RNA was stored under ethanol at -80°C.

DNA : DNA solution reassociation and hydroxyapatite chromatography

Driver DNA was usually sonicated to a length of 200-600 bp, (2 minutes on a Kerrys KT 100 sonicator). Labelled tracer DNA and at least a 5000 fold excess of driver DNA were heat denatured and reassociated in 0.12M phosphate buffer (1M PB is 0.5M Na dihydrogen orthophosphate, 0.5M disodium orthophosphate) at 60°C in sealed tubes. The period of incubation for the required Cot value was determined from the relationship

$$\text{Cot} = \frac{0.0260}{2} \times t \text{ (hours)}$$

After reassociation to the required Cot, single and double stranded DNA were separated on hydroxyapatite (HAP). Double stranded DNA was bound to HAP (Biorad. DNA grade) which had been equilibrated in 0.12M PB in water jacketted columns at 60°C. Single stranded DNA was eluted with 4 x 2ml rinses of 0.12M PB, the remaining double stranded DNA with 4 x 2ml rinses of 0.5M PB.

In thermal elution experiments, reassociated DNA was bound to HAP at 60°C and the temperature of the water jacketted column raised at intervals. Single-stranded DNA was eluted at each stage by 4 x 2ml rinses with 0.12M PB, TCA precipitated and

subjected to scintillation counting.

#### TCA precipitation and Scintillation counting

To each DNA sample to be TCA precipitated was added BSA to a final concentration of 50-100  $\mu\text{g/ml}$  and 20% TCA to a final concentration of 10%. After mixing this was kept on ice for 15 minutes before being filtered under suction through filter paper (Whatman GF/C) which had been soaked in 5% TCA. The filter was washed with 5% TCA before being dried for 1 hour at  $70^{\circ}\text{C}$  prior to scintillation counting.

In instances where small aliquots, e.g. 5 $\mu\text{l}$ , of labelled-DNA solutions were to be counted, these were spotted directly onto Whatman GF/C paper.

Dry filters were immersed in a 24:1 toluene liquifluor (NEN) scintillation cocktail and counted on an Intertechnique SL30 liquid scintillation spectrometer.

#### S1-nuclease digestion

To the DNA solution was added four volumes of 0.3M NaCl, five volumes of S1 assay buffer (0.05M acetic acid, 0.05M sodium acetate, 5mM  $\beta$  mercapto-ethanol, 2mM zinc sulphate pH 4.5) and an appropriate number of units of S1 nuclease enzyme (Sigma) in accordance with the manufacturers recommendations. The mixture was incubated at  $37^{\circ}\text{C}$  for the appropriate length of time before being either loaded onto a HAP

column or ethanol precipitated.

Chromosome preparation and staining.

Actively growing root tips were excised from healthy bulbs growing in water and incubated at room temperature in 0.2% fresh colchicine with occasional agitation (to ensure aeration) for six hours. After washing in distilled water the tips were fixed in 3:1 ethanol:acetic acid overnight at room temperature after which tips were hydrolysed in preheated 1M HCl for eight minutes at 60°C, washed in water and stained by immersion in Feulgen stain for one hour. The terminal 1-2mm of the stained tips were excised, teased apart and squashed in 45% acetic acid. Chromosomes were viewed and photographed on a Vickers M17 microscope on Ilford FP4 film.



## CHAPTER 3

### THE GENUS ALLIUM

## THE GENUS ALLIUM

### (1) Species classification

The genus Allium, described by Stearn (1944) as a "vast almost unwieldy genus", is extremely diverse, containing 600 or more species which cover the whole Northern hemisphere. The majority of species are perennial monocotyledonous wild plants which possess the taste and odour characteristic of onions and garlic; a property which has resulted in the cultivation of several species since Egyptian times. The resulting variety of cultivars makes the task of Allium taxonomy (already difficult due to the diversity of structure and the combinations of characters) even more difficult.

Early taxonomic works by Clusius and Haller are recognised as being important in the formulation of Linnaeus's Species Plantarum of 1753 in which 31 species of Allium were described. Despite numerous attempts, notably by Don and Regel in 19, no classification yet proposed is considered totally adequate, although in more recent years cytological techniques have contributed greatly to an understanding of species relationships. Such work is taken into consideration by Traub (1968) for example in his tentative key for Allium species.

A classification of the Allium species of Europe and

the Middle East is presented in figure 3 ; the majority of this information was obtained from J. Shaw (pers.comm). Only the main subgenera and sections are shown as taxonomy beyond this level becomes somewhat confused. Although the nomenclature is different to some extent, this classification agrees with others available (for example Traub, 1968; Tutin, 1980) in so far as the grouping of species into common lines of descent is concerned.

Some general information pertaining to the species used in this investigation is presented below. Specific details relating to chromosome number, ploidy level, DNA content etc. are presented in table 1.

A large amount of information relating to the botany, cultivation and utilisation of Allium species is presented by Jones and Mann (1963).



TABLE 1

Species name	Common name	Basic chromosome number	2n	Ploidy	DNA content 2c (pg)
<u>A. sativum</u>	common-garlic	8	16	2	32.7
<u>A. porrum</u>	leek	8	32	4	24.1
<u>A. babingtonii</u>		8	48	6	
<u>A. vineale</u>	crow-garlic	8	32	4	
<u>A. sphaerocephalon</u>		8	16	2	
<u>A. schoenoprasum</u>	chive	8	16	2	15.6
<u>A. hierochuntium</u>	shallot	8	16	2	30.3
<u>A. cepa</u>	onion	8	16	2	33.5
<u>A. fistulosum</u>	Japanese bunching-onion	8	16	2	26.3
<u>A. stipitatum</u>		8	16	2	
<u>A. tuberosum</u>	Chinese chive	8	32	4	34.8
<u>A. neapolitanum</u>		7	14	2	
<u>A. moly</u>	Golden Allium	7	14	2	45.2
<u>A. ursinum</u>	Ramsons	7	14	2	63.0
<u>A. ostrowskianum</u>	Red-star	7	14	2	39.8/ 44.6
<u>A. triquetrum</u>		9	18	2	36.3

(Data from references cited in main text)

### Species Notes

#### SECTION ALLIUM

##### A. sativum L.

The common garlic A. sativum (sativum = cultivated) occurs in a variety of morphological types, all of which are completely seed sterile, reproduction therefore being exclusively vegetative. Consequently existing cultivars must have arisen as vegetative mutations during the long cultivation of garlic.

Three basic types of A. sativum, designated H, U and A, distinguishable on the basis of morphological characters, have been described (Konvicka and Levan, 1972). The chromosomes of A. sativum have been described by several workers (Verma and Mittal, 1978; Konvicka and Levan, 1972). All three types have 16 chromosomes which, although similar in gross morphology between clonal types, do show slight differences in length and arm ratios between H, U and A.

##### A. ampeloprasum L. (ampelos = grape vine, prason = leek)

Wild forms of this species are widespread in their distribution throughout Europe and are extremely variable. The cultivated forms can be divided into three groups, one of which, the 'leek group' contains the cultivated leek A. ampeloprasum var porrum (L) Gay, hereafter referred to simply as A. porrum.

A. babingtonii Borrer

Considered a variety of A. ampeloprasum by some authors, A. babingtonii is unusual in being unknown outside Britain. First noted in Cornwall it has since been reported in Dorset and Western Ireland.

A. sphaerocephalon L.

A very variable species throughout Europe with local differentiation.

A. vineale L. ('crow garlic')

A species widespread in most of Europe as a weed on cultivated land.

## SECTION CEPA

A. cepa L.

Unknown in the wild state, A. cepa is probably derived from A. oschanini in the region of Iran and has been in cultivation for over 3000 years. The cultivars can be divided into three groups, the majority, (including 'spring onion' 'tree onion' and A. cepa, used in this study) belonging to the 'common-onion' group.

The shallot, A. hierochuntium (or A. ascalonicum) belongs to the 'aggregatum' group.

A. fistulosum L.

The Japanese bunching onion (or Welsh onion) is

unknown as a wild species, having many cultivated forms. It has been crossed successfully with A. cepa to produce a hybrid.

#### SECTION RHIZERIDEUM

##### A. tuberosum

The chinese chive is known as both a wild and a cultivated plant. It is unusual in that it spreads by rhizomes rather than by forming bulbs.



## (2) The chromosomes and DNA of Allium species

### i) Chromosomes

By 1965 about 170 forms of Allium had been studied cytologically, revealing there to be three basic haploid complements;  $n = 7$ ,  $n = 8$  and  $n = 9$ , the majority having  $n = 8$  or a multiple thereof since polyploidy is common and ranges up to 13x. Despite this fact, it seems to be generally accepted that the primary basic chromosome number in the genus Allium, from which the others have been derived, is  $n = 7$ . It is thought that the present day Karyotypes and distribution of the species were largely determined by a drastic reduction in the number of plants during the last glacial period and the subsequent reexpansion of populations of survivors afterwards, an example of a genetic bottleneck.

Confirmation of the somatic complement number of several Allium species was sought by subjecting metaphase chromosomes of squashed root tip material to Feulgen staining (see Materials and Methods).

Typical examples of stained metaphase chromosomes of A. sativum are shown in figure 4 confirming the somatic chromosome complement to comprise 16 chromosomes, i.e. a diploid in which  $n = 8$ . Although accurate Karyotyping is difficult using only Feulgen stained material, the eight pairs of chromosomes can be divided into three groups as described by Konvicka and Levan (1972)

and Verma and Mittal (1978):

- 1) 5 long pairs, each with approximately median centromeres
- 2) 2 satellited pairs, both of which are more asymmetric than the 'long' pairs and which have secondary constrictions separating a short proximal segment from a big satellite on their short arms
- 3) 1 short pair, distinguishable by size, with clearly asymmetric arms

The results of squashes of material from A. babingtonii ( $2n = 48$ ), A. cepa ( $2n = 16$ ), A. triquetrum ( $2n = 18$ ), and A. porrum ( $2n = 32$ ) are shown in figure 5. In all cases the complement number presented in table 1 is confirmed. The variability of both chromosome number and chromosome volume between species within the genus Allium is clearly demonstrated.

Supernum<sup>er</sup>ary B-chromosomes have been reported in several Allium species (Ved Bra't, 1965a). A clinal variation in B-frequency has been detected in natural populations of A. schoenoprasum (Bougourd and Parker, 1975) while a correlation between the frequency of B-chromosomes and germination time has been reported in A. porrum (Vosa, 1966).

FIGURE 4

Metaphase chromosomes of A. sativum.

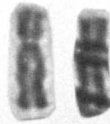
(a) metaphase squash of the somatic complement

(b) idiogram



A

10μm



B

10μm

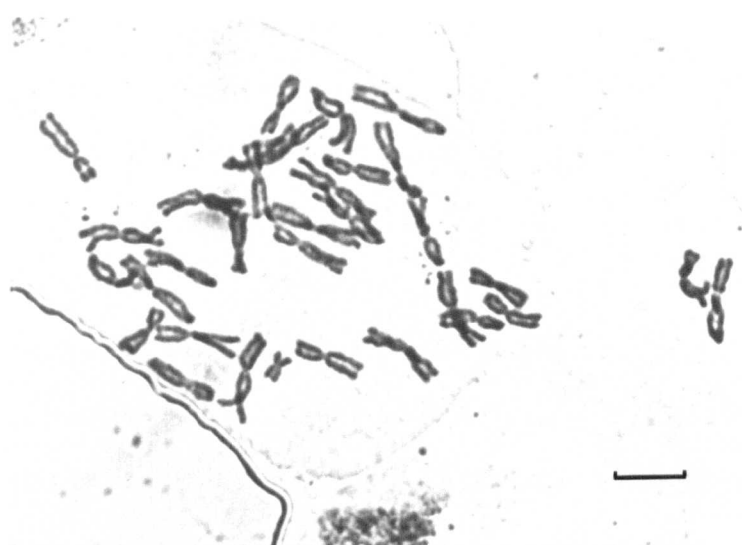
FIGURE 5

Metaphase chromosomes of A. cepa (a),  
A. porrum (b), A. babingtonii (c) and  
A. triquetrum (d), (not to scale).

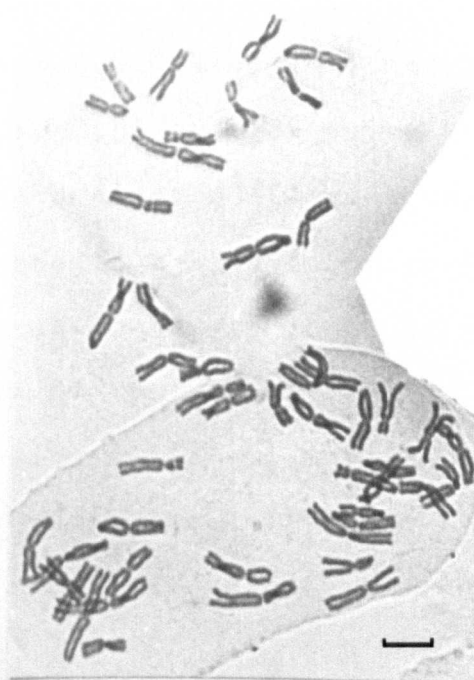
— represents 10  $\mu$ m



a



b



c



d

## ii) DNA

A variation in nuclear DNA content between members of different Allium species was perhaps first observed by Levan (1935) in describing a 3-fold range of chromosome lengths in various species. Jones and Rees (1969) have more recently demonstrated significant differences in nuclear DNA content between diploid species of each basic chromosome number, constancy of chromosome number clearly not reflecting constancy of DNA amount. DNA amount however is, not surprisingly perhaps, correlated with chromosome volume. Because of the poorly understood taxonomy, they conclude that although changes in DNA amount are widely distributed between (but may be highly localised within) chromosomes, the evolutionary direction of such changes cannot be correlated with the phylogeny of Allium species.

The hope that an investigation into the base composition of Allium nuclear DNA would prove useful as a taxonomic tool was dashed by the resulting range of 34.6 - 39% G + C being too narrow (Kirk et al., 1970). Furthermore the lack of correlation between base composition and nuclear DNA amount suggests that there is no base composition bias in those DNA sequences which can be observed to have changed in quantity during the evolution of the genus.

Despite the above work and a multitude of papers concerned with the Karyotypic investigations of Allium species (for example Konvicka and Levan, 1972; Verma and

Mittal, 1978), there are relatively few studies concerned with the physical organisation of DNA in such species.

Perhaps the most significant to date is the study of the chromosomes and DNA of A. cepa (Stack and Comings, 1979). A Cot curve provided evidence for there being three fractions of the genome. The estimated proportion of the genome made up by repetitive DNA is one of the highest for any plant species so far investigated, while interestingly the estimated value of G + C content, one of 33.2%, is the lowest so far reported for an angiosperm. In situ hybridisation suggested that moderately repetitive sequences were distributed throughout the length of all A. cepa chromosomes while evidence for interspersion of repetitive and single copy DNA sequences at the molecular level, from work on the reassociation of fragments of different lengths, indicated the A. cepa genome to be predominantly of the short period interspersion pattern.

Cot curves for 7 Allium species have been prepared by Ranjekar et al. (1978) and have been interpreted as suggesting that differences in the genome size among those species studied are not due exclusively to variations in the percentage of repetitive DNA. Repetitive fractions were further characterised by isopycnic centrifugation and thermal denaturation, demonstrating a gross similarity between species.



Several Allium species have been alluded to in investigations concerned with the search for satellite DNA in plant genomes. In no instance was the presence of any measurable quantity of satellite DNA reported (Ingle et al., 1973; Ingle et al., 1975; Stack and Comings, 1979). In contrast however, a tandemly repeated 400 bp sequence has been identified and cloned from the genome of A. vineale (data presented later). A similarly sized and apparently related satellite component has also been identified and characterized in the genome of A. cepa (A. James and S. Barnes, pers. comm.). Sequences closely related to these highly repetitive tandemly organised sequences of A. vineale and A. cepa are also present in the genomes of several other Allium species.

A Cot curve for the nuclear DNA of A. sativum has been generated (Evans et al., in press) and agrees well with that presented by Ranjekar et al. (1978) for the same species. There appear to be three major components of the genome. Approximately 25% of the genome reassociates before  $\text{Cot } 5 \times 10^{-2}$ , presumably including the very highly repetitive sequences, (such as satellite if it is present) and 'fold back' DNA. A fraction representing approximately 40% of the genome, reassociating between  $\text{Cot } 10^{-1}$  and  $\text{Cot } 10$  represents the 'moderately repetitive' component of the genome while

single-copy and low-copy sequences are presumably present in the fraction reassociating after  $Cot\ 10^2$ , a fraction which will also include any unreactable fragments.

Further investigations into the general organisation of the nuclear genome of A. sativum, A. cepa and A. porrum have employed DNA:DNA solution reassociation techniques to construct interspersion curves. Tracer DNA fragments of different lengths from one species were reassociated with an excess of driver DNA from the same and from different species (Evans et al., in press; Evans, James and Barnes, unpublished). The results suggest that many repetitive sequences are common to the genomes of all three species. Such shared sequences seem to make up different proportions of each genome and to show some sequence divergence between genomes. Their physical organisation however appears to be approximately the same in each species, many of the repetitive elements being interspersed with sequences of lower repetition.

## CHAPTER 4

### THE CHARACTERISATION OF B1000 IN THE A. SATIVUM GENOME

## THE CHARACTERISATION OF B1000 IN THE

### A. SATIVUM GENOME

#### (i) Restriction endonuclease digestion of

##### A. sativum DNA

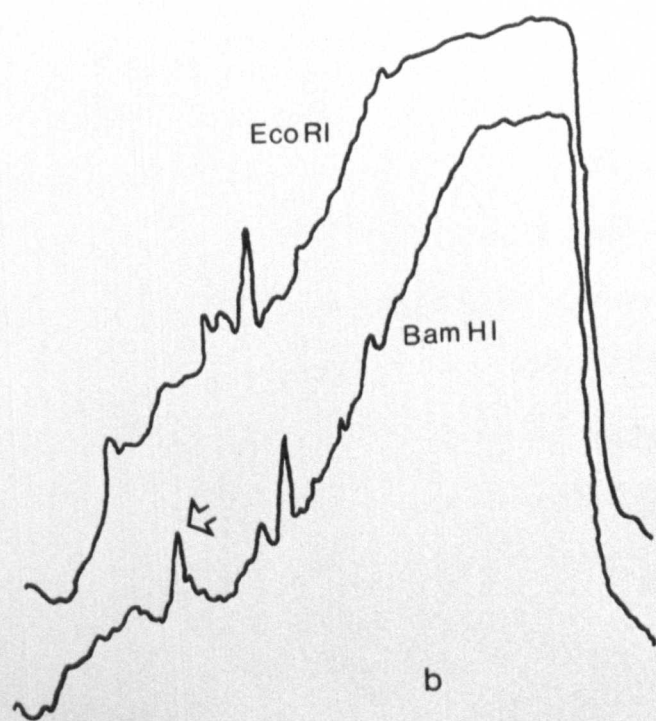
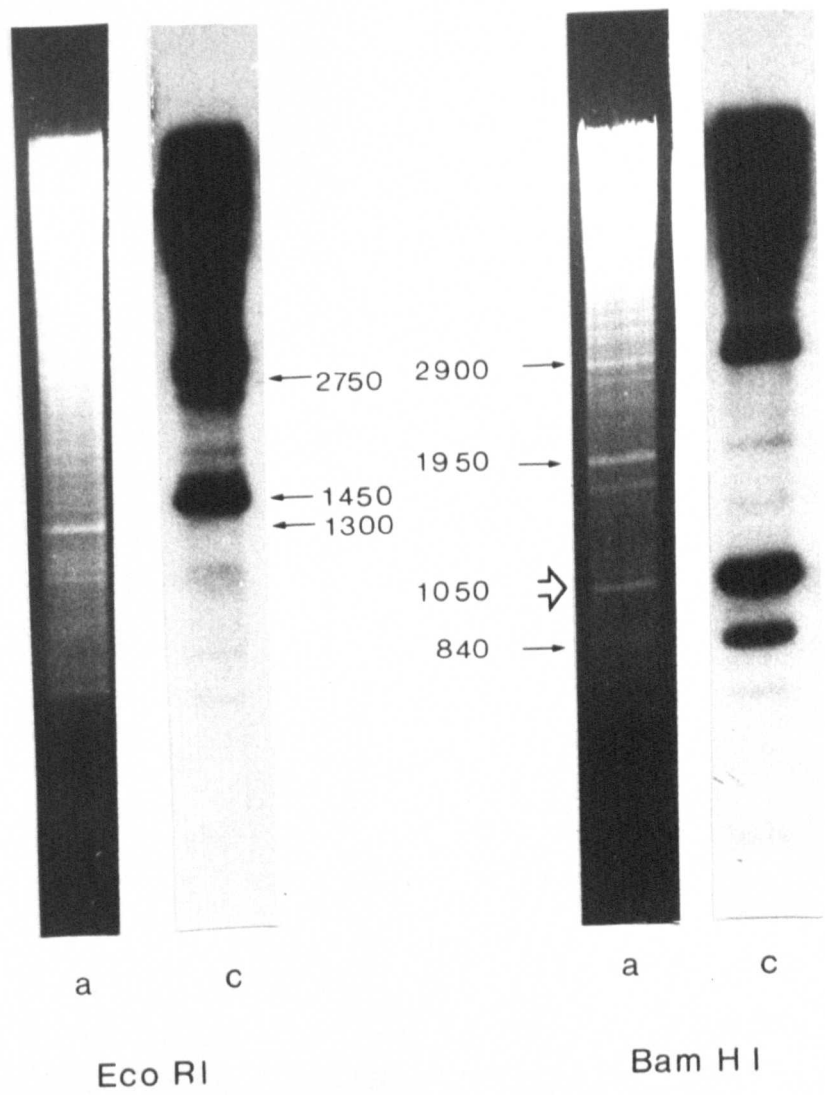
Total genomic DNA from A. sativum is rendered susceptible to type II restriction endonucleases after removal of associated contaminating molecules such as proteins, polysaccharides and RNA. This was best achieved by CsCl density gradient centrifugation in the presence of ethidium bromide as described in Materials and Methods, DNA extracted by this method typically possessing an OD 260/OD 280 ratio of approximately 1.4.

A number of restriction endonucleases, having 4 and 6 base-pair recognition sites, cleave total A. sativum DNA into fragments which, when fractionated by gel electrophoresis and stained with ethidium bromide appear as discrete bands over a background smear. Such bands presumably comprise repeated sequences with conserved restriction sites while the background smear represents the fractionation of fragments liberated by scission of the DNA at effectively randomly distributed sites, as might be expected in the case of single copy sequences or of very diverged repetitive sequences.

Typical results are shown in figure 6(a). Total DNA from A. sativum was digested separately

FIGURE 6

- a) A. sativum DNA digested with Eco RI and Bam HI, fractionated on an agarose gel and stained with ethidium bromide.
  - b) Densitometer scans of the digests shown in figure 6(a).
  - c) Autoradiographs of the digests shown in figure 6(a) after Southern blotting and hybridisation with  $^{32}\text{P}$  B1000.
- The sizes of some of the major bands are marked in base pairs. The band representing the source of the B1000 probe is arrowed.



with Eco RI and Bam HI restriction endonucleases, electrophoresed on a 1% agarose gel, stained with ethidium bromide and photographed with UV illumination. A number of strong bands are apparent, the size of many of which are indicated, ranging on this gel from approximately 600 bp to over 8000 bp. The size distribution of fragments is displayed in a different form in figure 6(b), where the photographs of figure 6(a) have been scanned in a densitometer.

The patterns of digestion are characteristically different, very few bands being common to the two digests.

This is indeed the situation in fractionated digests using most other restriction endonucleases (data not shown), immediately suggesting that most of the DNA in the bands is not organised in the genome in long tandem arrays. Typically in such an organisation, the same sized repetitive fragments may be released by restriction endonucleases recognising different sites (see for example Chapter 7). Furthermore there is no evidence of the ladder of fragments typical of satellite DNA in any of the restriction digests of A. sativum so far observed. Nevertheless it is possible that the complexity of the genome might hide a minor component with such characteristics.

It is practically impossible to interpret any higher-order sequence arrangements from ethidium bromide stained electrophoresed restriction fragments of A. sativum DNA (even using combinations of restriction enzymes) simply because of the large size and complexity of the genome. There are clearly a large number of fairly homogenous repetitive elements, as evidenced by the number and distribution of bands, but their relationship to one another cannot be determined from data such as that in figure 6(a), particularly as the prominent bands in each digest display no obvious numerical relationship to each other.

For this reason it was necessary to simplify the investigation by focussing attention upon the organisation of one family from the complex population of genomic sequences.

(ii) The cloning of a repeated DNA sequence from the A. sativum genome

Of the many prominent bands superimposed upon the fluorescence of DNA in figure 6(a), that of 1050 base pairs in the Bam HI digest was considered the most favourable for further investigations for a number of reasons;

1) The DNA in the band makes up a small but fairly significant proportion of the genomic DNA.

Densitometric scanning of the Bam HI digest



(figure 6(b)) suggests that the band represents approximately 0.6% of the total.

2) The fragments making up the band are reasonably short compared to the majority of other bands and therefore easier to manipulate. Problems in cloning procedures have been reported when using large fragments, particularly those which are internally repetitious and consequently impart instability (Bedbrook & Gerlach, 1980).

3) It is well resolved from the bulk of restriction fragments on a gel, aiding the ease of isolation and minimising contamination by co-migrating DNA.

Total A. sativum DNA was digested with Bam HI, fractionated on a 1% agarose gel, stained with ethidium bromide and a strip of gel containing the band of 1050 bp excised. The DNA contained within was isolated by electroelution and labelled with  $\alpha$  <sup>32</sup>P - dCTP by nicktranslation. This probe was used to screen a small partial library of Bam HI fragments of A. sativum inserted into pBR322 plasmid vector. All these procedures are described in Materials and Methods.

Three clones (from 65) were picked out in this way and one of these, pAs 3017 was found to contain a Bam HI fragment 1050 bp in length and homologous to the probe. This fragment was

subcloned into pAT 153 vector and this clone, pAs 3017B, used as a source of the repetitive fragment hereafter referred to as the B1000 sequence (Ba<sub>m</sub> HI, approx. 1000 bp).

(iii) B1000 sequence is found in a number of genomic configurations

The availability of pure B1000 sequence in large quantity enabled its organisations in the A. sativum genome to be investigated by virtue of the fact that under suitable conditions radioactively labelled B1000 DNA will hybridise to filter-bound fragments sharing some homology. The degree of hybridisation depends on a number of factors such as the stringency of reassociation, the degree of homology of the reacting molecules etc., factors discussed in depth by Britten et al. (1974) for solution reassociation experiments.

Total A. sativum DNA digested with Eco RI and Ba<sub>m</sub> HI and fractionated on a 1% agarose gel (that presented in figure 6(a)) was transferred by Southern blotting onto nitrocellulose filter and hybridised with B1000 probe as described in Materials and Methods. The resulting autoradiographs are presented in figure 6(c) where they are compared to the fractionated DNA prior to blotting. In addition to hybridisation to fragments of 1050 bp

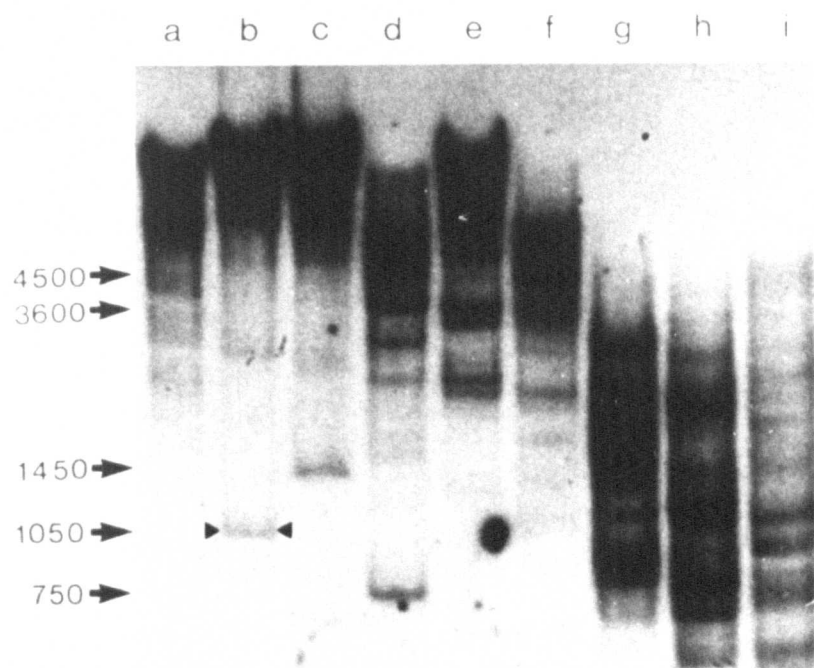
in a Bam HI digest, (the source of the B1000 probe), there is considerable hybridisation to fragments whose lengths show no obvious relationship to that of the probe, or indeed to each other. Not only is there hybridisation, of varying degrees of intensity, to other discrete bands both larger and smaller than the probe length, but also to a background smear. The fact that there are bands larger than 1000 bp in addition to a smear suggests that if B1000 is interspersed in the genome, it is not entirely diffuse, some copies being non-randomly arranged in larger repeating units. The same conclusion can be drawn from the distribution of hybridisation to the Eco RI digested DNA. As an extension of this experiment, total A. sativum DNA was digested with a number of different restriction endonucleases and probed with B1000 as described above. The pattern of hybridisation was characteristic for each restriction enzyme, examples of which are shown in figure 7(a), there being pronounced discrete bands of hybridisation over a background smear. In no digest is a satellite-ladder apparent, nor do the prominent bands bear any simple numerical relationship to one another within or between digests. The situation is clearly too complex to permit even the construction of tentative restriction maps

FIGURE 7

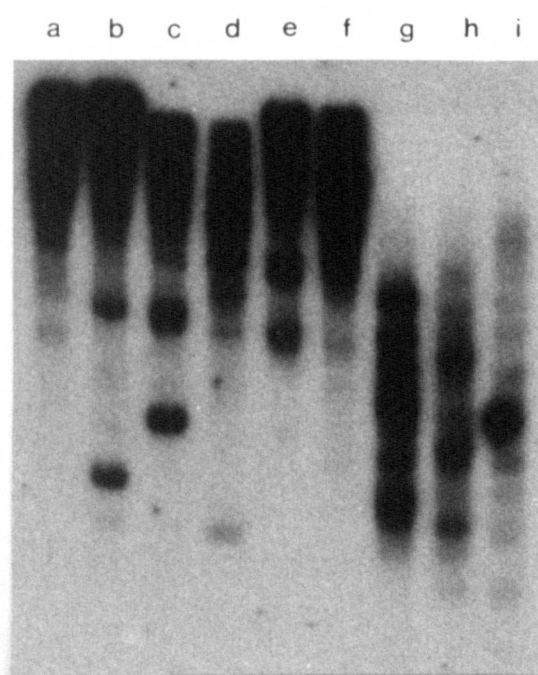
- a) Autoradiograph of a Southern blot of A. sativum DNA digested with (a) Hind III (b) Bam HI, (c) Eco RI, (d) Bcl I, (e) Bql II, (f) Hinc II, (g) Hae III, (h) Alu I and (i) Hinf I and hybridised with  $^{32}\text{P}$  B1000.

Hybridisation of the B1000 probe to its position of origin in a Bam HI digest is indicated by (► - ◄); the size of some other major bands is marked in base pairs.

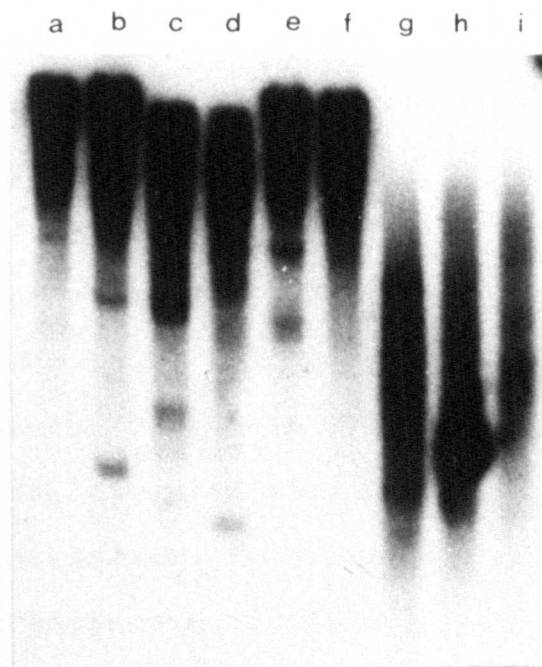
- b) Autoradiograph of cloned B1000 probe hybridised to A. sativum DNA digested with (a) Hind III, (b) Bam HI, (c) Eco RI, (d) Bcl I, (e) Bql II, (f) Hinc II, (g) Hae III, (h) Alu I and (i) Hinf I.
- c) The Southern blot of figure 7(b) reprobbed with  $^{32}\text{P}$  genomic 1050 bp-band DNA.



a



b



c

of elements containing B1000 homologous sequences.

(iv) The cloned B1000 sequence is a representative member of the genomic sequence family

Although B1000 hybridises strongly to the 1050 bp band in a Bam HI digest of A. sativum it could have been formally possible that the probe was not typical of the majority of sequences in that band.

Therefore a comparison was made between the hybridisation patterns resulting from the use of cloned B1000 and of the genomic 1050 bp band excised from a gel as probes against total DNA digested with several different restriction endonucleases. The latter probe represented all variants of sequences of length 1050 bp in proportion to their numerical frequency in the band whereas B1000 probe represented just one sequence which may or may not be typical. (B1000 was however originally isolated by its hybridisation to the nick-translated genomic band in the initial screening and ought therefore to be fairly homologous to many of the sequences contained therein).

Total A. sativum DNA was digested with several different restriction endonucleases, fractionated on a 1% agarose gel and transferred to nitrocellulose paper. Initially this Southern blot was probed with nick-translated cloned B1000 sequence, resulting in the hybridisation pattern shown in

figure 7(b).

After sufficient time for the  $^{32}\text{P}$  of the probe to decay below detection level (rather than eluting the probe and risking washing off some of the filter bound driver DNA) the same filter was re-probed with  $^{32}\text{P}$  labelled genomic 1050 bp band DNA, resulting in the hybridisation pattern shown in figure 7(c).

The distribution of hybridisation on each autoradiograph is very similar. The same major bands are highlighted although there is more hybridisation to a background smear in the case of the genomic 1050 bp band probe, particularly in the digests using enzymes with 4 bp recognition sites. This would be expected if there is a small degree of sequence contamination by non-B1000 sequences in this more complex of the two probes or indeed if there is sequence heterogeneity of B1000-like sequences within the genomic band.

The fact that the two autoradiographs are essentially identical would seem to suggest that the cloned B1000 sequence is indeed typical of the average or most common sequence in the genomic 1050 bp band.

(v) Estimation of the Repetition frequency of the B1000 sequence in the *A. sativum* genome

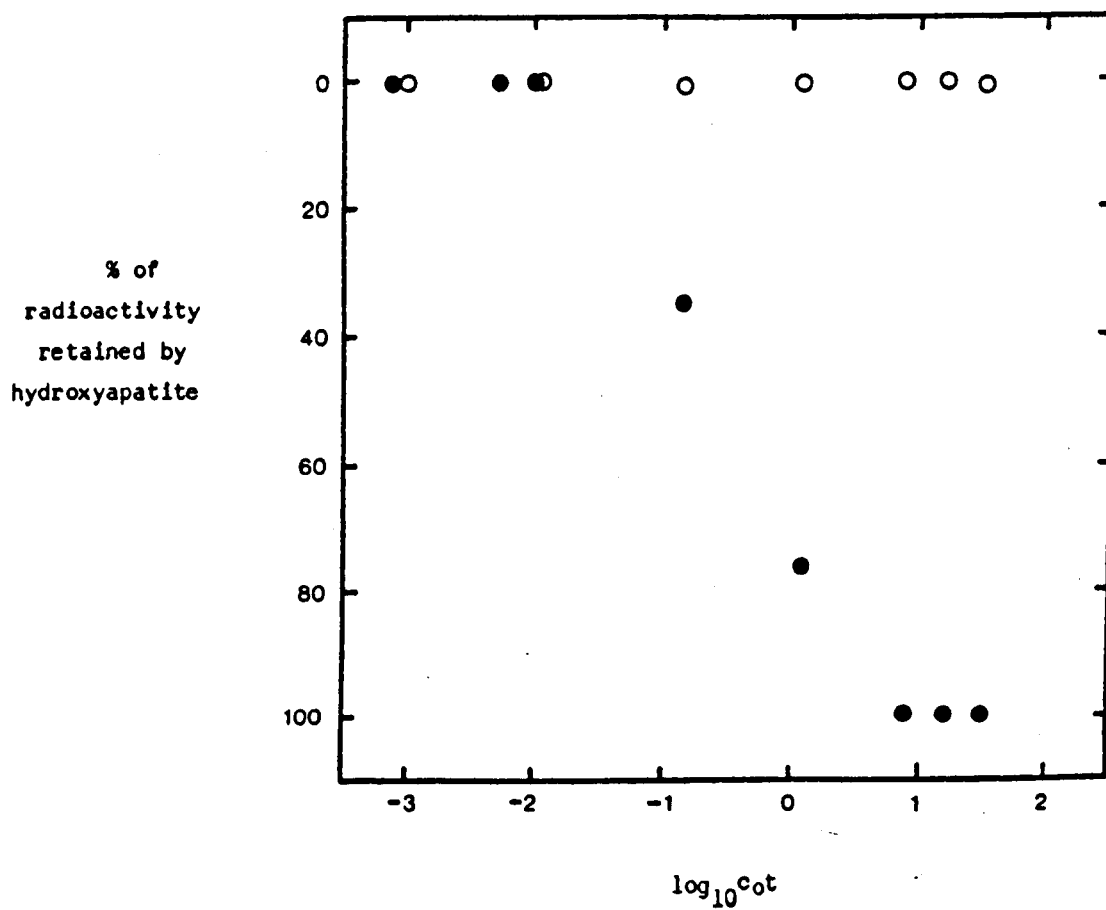
The amount and range of hybridisation of B1000 probe to genomic blots (see for example, figure 7 above) suggests that the probe is of high reiteration

frequency in the A. sativum genome. Confirmation of this fact was sought by monitoring the solution reassociation kinetics of  $^{32}\text{P}$  labelled B1000 used as a tracer against total A. sativum driver DNA. Reassociation was monitored by hydroxyapatite column fractionation of single and duplex molecules as described in Materials and Methods.

Figure 8 shows the resulting reassociation curve. Reassociation of B1000 probe using excess calf thymus DNA as a driver was monitored in order to detect any self-reassociation during the term of the experiment. None was detectable by Cot 50, suggesting B1000 contains no fold-back sequences. Driven by A. sativum DNA, B1000 probe reassociated between Cots  $10^{-2}$  and 10, typical of sequences in the moderately repetitive component of the Cot curve of A. sativum DNA.

From the  $\text{Cot}_{\frac{1}{2}}$  value of its reassociation, the copy number or the complexity of the B1000 homologous sequences can be estimated (Lewin, 1980). The  $\text{Cot}_{\frac{1}{2}}$  from figure 8 equals  $5 \times 10^{-1}$ . Assuming a repeat length of approximately 1000 bp the reassociation kinetics are consistent with a copy number of approximately  $3 \times 10^4$  although, as with many reassociation studies there are inaccuracies inherent in this calculation, the kinetics of reassociation being particularly affected for example by any





**FIGURE 8**

Reassociation of Bl000 driven by A. sativum total DNA (● - ●) and calf-thymus DNA (○ - ○).

sequence heterogeneity.

An estimate of copy number can also be obtained from a comparison between the fluorescence of DNA on ethidium bromide stained gels and the degree of hybridisation of B1000 probe to the same DNA transferred to nitrocellulose filter.

The 1050 bp band appears to make up approximately 0.6% of the total genomic DNA when A.sativum DNA is digested with Bam HI and fractionated on a gel, a figure determined earlier from densitometric data (see figure 6(b)). However, as demonstrated by the hybridisation patterns in figure 7 for example, B1000-homologous sequences are present on many other fragments in such a gel. In fact the band of 1050 bp seems to comprise only approximately 6.8% of the total radiotope-induced darkening on a typical autoradiograph, as determined by microdensitometry.

Consequently sequences homologous to the B1000 probe, at the stringencies applied in the hybridisation experiments, make up approximately  $100/6.8 \times 0.6\%$  of the genome or 8.8%. Assuming again a repeat length of 1000 bp, and a genome size of  $3 \times 10^{10}$  bp (calculated from a 2c value of 32.7 pg; see table 1), such a value corresponds to approximately  $2.5 \times 10^6$  copies per genome ( $8.8\% \times 3 \times 10^{10}/1000$ ). This calculation assumes

that all the sequences in the genomic 1050 bp band will hybridise to B1000 probe i.e. that they are all related. In this respect the copy number estimate is a maximum value.

Considering the inaccuracies inherent in both these methods of copy number estimation, the figures of  $3 \times 10^4$  and  $2.5 \times 10^6$  are not too dissimilar.

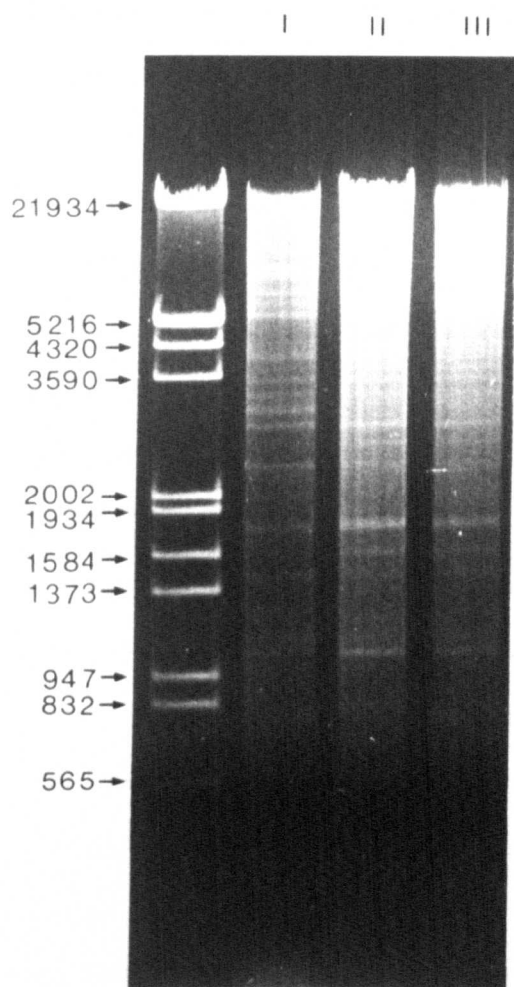
(vi) The copy number and physical organisation of B1000 does not appear to be tissue specific

To investigate whether either the reiteration frequency or the organisation of B1000-like sequences differs between tissues of the A. sativum plant,  $^{32}\text{P}$  B1000 was hybridised to nitrocellulose-bound Bam HI digested DNA extracted from three different regions of a week-old individual plant. Total DNA was extracted from root, bulb and shoot tissues as described in Materials and Methods but on a smaller scale and without density gradient purification. No attempt was made to eliminate plastid DNA such as mitochondrial or chloroplast DNA.

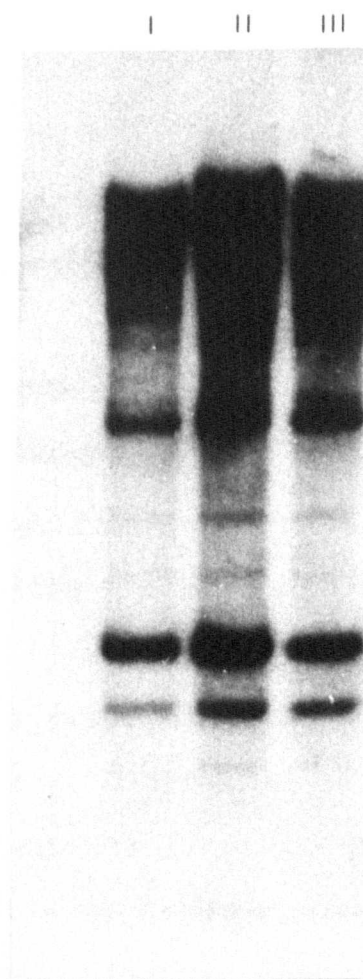
The results of digestion of DNA from each source with Bam HI fractionated on a 1% gel are shown in figure 9(a). Digestion, while not entirely complete in all cases, compares very favourably in its extent with the digestion of DNA purified on a larger scale by CsCl centrifugation (see figure 6(a)). The

FIGURE 9

- a) DNA extracted from i) shoot, ii) bulb and iii) root tissue of A. sativum, restricted with Bam HI and fractionated on an agarose gel.
- The sizes of the  $\lambda$ Eco RI/Hind III digest used as a marker are given in base-pairs.
- b) Autoradiograph of the DNA of figure 9(a) after Southern blotting and hybridisation with  $^{32}\text{P}$  B1000.



a



b

positions and intensities of major bands are clearly very similar in digested DNA from root, bulb and shoot tissue, being practically identical to the pattern presented in figure 6(a) for Bam HI digested gradient purified bulb DNA.

The autoradiograph produced when the DNA from this gel was transferred to nitrocellulose and hybridised with B1000 is presented in figure 9(b). Taking the slightly different concentrations of each DNA on the gel into consideration, the degree of hybridisation to each appears to be very similar. Likewise the pattern of hybridisation appears to be the same, very closely resembling that presented earlier in figure 6(c).

It would appear from this result therefore that the frequency and organisation of B1000-like sequences is the same throughout the plant, at least at the level of discrimination possible by crude extraction of DNA from root, bulb and shoot tissue; there is no apparent significant somatic amplification or rearrangement.

Of course each of these three tissue types themselves contain a wide variety of different types of cells at different developmental stages, and it is possible that there is some variation in the frequency and organisation of B1000 sequences at a lower level in the plant.

Changes in the DNA content of cells as a result of differential replication of components of the genome have been observed during plant development (Nagl, 1982). The phenomenon of endo-reduplication has been described in A. cepa in which there appear to be differences in re-association kinetics between DNA extracted from different segments of root tissue (Cremonini et al., 1981). This has been interpreted as being evidence for differential replication of highly and moderately repetitive sequences in certain cell-types.

(vii) The length of the B1000-associated repetitive element

Two situations can be envisaged to explain the patterns of hybridisation such as those presented in figures 6 and 7, where strong bands (mostly larger but some smaller than the probe length) are highlighted by B1000 probe.

The B1000 could be part of a larger repeating unit in which case the distinctive hybridisation patterns would be due to heterogeneity of restriction sites within and external to the larger repeat, perhaps indicative of such a larger element being found in a number of different sequence environments within the genome. Alternatively the complex pattern of bands could be due to the B1000 sequence being composed of a number of smaller repetitive elements which are themselves found in a number of different

permutations within the genome.

Thus in both of the above cases, a complex hybridisation pattern of discrete bands superimposed upon a background smear is due to the same phenomenon i.e. a combination of restriction site heterogeneity and repetitive element sequence permutation. The two cases differ in their scale. In the former the B1000 is part of a larger repetitive element, whereas in the latter it is composed of a number of smaller elements.

One way to discriminate between these alternatives is to determine the hybridisation patterns produced by using different parts of the B1000 sequence as probes against digested genomic DNA. If B1000 is part of a larger repetitive element, all parts should behave in the same way as the whole in hybridisation to filter bound digested DNA (except when being used to probe fragments produced by enzymes recognising sites within the B1000-like sequence).

Conversely if B1000-like sequences are composed of a number of different unrelated repetitive elements, sub-fragments of the cloned probe might be expected to behave differently than the whole, each presumably picking out a subset of the bands produced by the whole.

In order to generate sub-fragments of the cloned B1000, restriction enzymes recognising sites within



the sequence were sought. Three enzymes were found to cut B1000 into sub-fragments of significant size; all three, Hae III, Alu I and Hinf I having 4 base pair recognition sites. A detailed restriction map of B1000 showing the positions of Hae III and Alu I restriction sites is presented later (figure 13).

Digestion of B1000 with Hae III produces two fragments of approximately 610 and 440 bp. These fragments were isolated from an agarose gel, nick-translated and used separately to probe Southern blots of genomic A. sativum DNA digested with several combinations of enzymes.

The resulting autoradiographs are presented in figure 10. Clearly the hybridisation patterns are very similar for both probes. This suggests, but does not conclusively prove, that in the majority of instances the two Hae III fragments of B1000 are found contiguously arranged in the genome, presumably as co-linear parts of a larger repetitive unit.

If this were indeed the case it might be argued that the characteristic pattern produced by the hybridisation of B1000 to for example a Bam HI digest of A. sativum DNA is due to the partial digestion of a larger repeating structure. Although the hybridisation patterns are very reproducible and in most

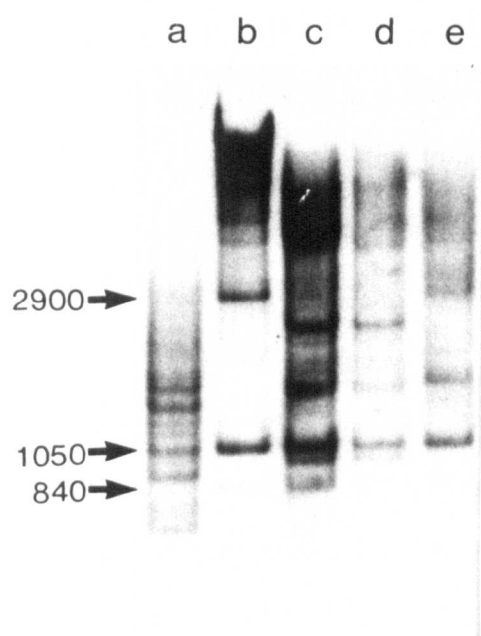
FIGURE 10

A. sativum DNA digested with (a) Bam HI/Hae III, (b) Bam HI, (c) Bam HI/Eco RI/Hind III, (d) Bam HI/Eco RI and (e) Bam HI/Hind III and probed with (I) the 610bp Hae III fragment and (II) the 440 bp Hae III fragment.

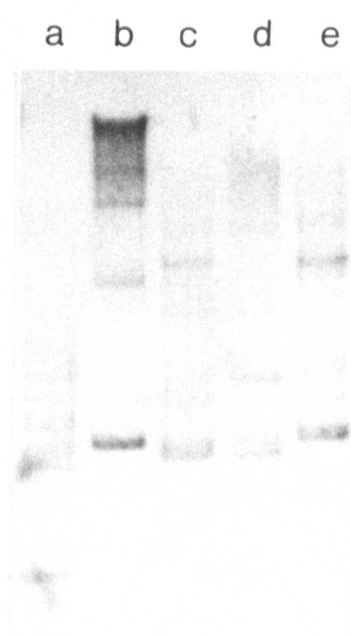
FIGURE 11

Time course Bam HI digestion of A. sativum DNA hybridised with  $^{32}\text{P}$  B1000.

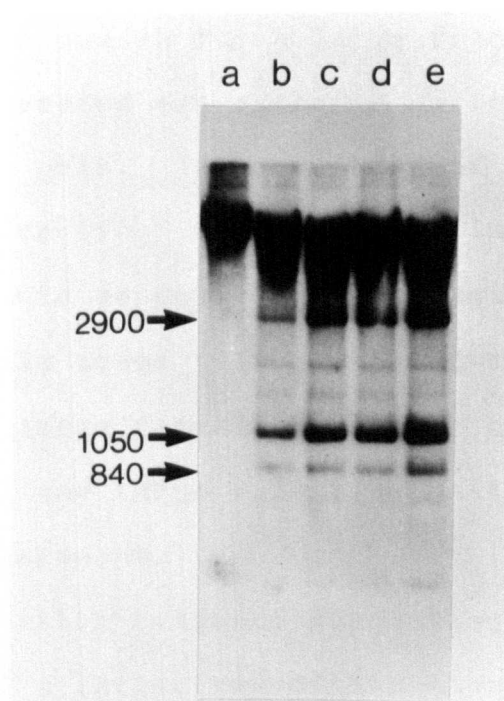
Lane (a) undigested DNA, (b) DNA digested for five minutes with 5 units enzyme, (c) 15 minutes with 5 units, (d) 45 minutes with 5 units and (e) 120 minutes with 20 units.



I



II



digests a vast excess of enzyme is used, this possibility was investigated by probing, with B1000, genomic DNA samples subjected to increasing periods of digestion with Bam HI.

A depletion of larger bands in concert with the appearance of other smaller ones with increasing degrees of digestion would suggest that many bands were indeed due to partial digestion of a large repeat unit. However, as figure 11 demonstrates, there appears to be no such depletion/appearance effect; all discrete bands in the displayed range increase in intensity with increasing digestion time.

This suggests that all these fragments are liberated independently from their sequence environment, and none are the result of incomplete digestion of others.

The search for a large fragment from which others are liberated was extended up to 15 Kb on weaker agarose gels. The same kind of results as presented in figure 11 were obtained in that no 'parent' band could be detected (data not shown).

This seems to suggest either the existence of a very large repeating unit or heterogeneity in the size of any large repeating unit(s) of which B1000 is a component.

Further evidence suggesting that B1000 is indeed part of a larger repetitive element is provided

by experiments involving reassociated repetitive DNA.

Mildly sonicated A. sativum DNA was reassociated to Cot 12 and digested with S1 nuclease. The S1-resistant duplexes were fractionated on an agarose gel, transferred to nitrocellulose filter and hybridised with  $^{32}\text{P}$  B1000. Hybridisation was almost entirely to DNA fragments of length greater than 7 Kb (data not shown). No discrete bands of hybridisation were however visible, possibly suggesting heterogeneity in the length of any repetitive element. A. sativum DNA which had been reassociated and S1-digested under the same conditions was fractionated on a Sepharose 2B column. Selected size fractions were run out on 2% and 0.5% agarose gels, transferred to nitrocellulose filter and hybridised with  $^{32}\text{P}$  B1000. Hybridisation was to a wide range of fragments, there being again no bands suggestive of discrete-length duplex molecules (data not shown).

The degree and effects of hyperpolymer formation (Flavell and Smith, 1977) in these reassociation experiments were however undetermined. Hybridisation of B1000 to duplexes greater than 7Kb in length can thus only be regarded as suggestive rather than conclusive, since incomplete removal of such hyperpolymer molecules by S1 nuclease digestion would confuse the interpretation of the above data.

(viii) B1000 sequences are not abundantly transcribed

Transcription of moderately repetitive DNA sequences has been demonstrated by hybridisation of specific probes to filter bound RNA - so called 'Northern' blots (see for example Posakony et al., 1983).

RNA was extracted from fresh root tissue as described in Materials and Methods. When fractionated on a 1% agarose gel and stained with ethidium bromide, ribosomal RNA bands were clearly visible, demonstrating that degradation during extraction and fractionation was minimal (data not shown). The rRNA bands were still visible under UV illumination after transfer of nucleic acids to nitrocellulose paper, before baking and subsequent hybridisation with  $^{32}\text{P}$  B1000. Hybridisation was performed both at  $65^{\circ}\text{C}$ . in the conditions employed for Southern blots as described in Materials and Methods and also at  $42^{\circ}\text{C}$  in a similar hybridisation solution containing 50% deionized-formamide.

In many experiments no hybridisation of B1000 to the filter was observed. Occasionally however some hybridisation was apparent (after very long exposure times) in the form of a smear.

Despite treatment of samples with RNase and

DNase it was not possible to clearly establish whether hybridisation was to RNA or to a minor component of sheared DNA. Hybridisation in 50% formamide at 42°C. favours DNA:RNA reassociation but in instances when hybridisation was detectable it was usually to the DNase treated sample alone. Hybridisation of probe to incompletely digested DNA cannot be discounted although as DNA was not denatured prior to transfer to nitrocellulose filter it ought not to be in hybridisable single stranded form on the filter.

The results of these experiments are thus unfortunately inconclusive. It would appear though that if B1000 sequences are present on RNA molecules that their transcription rate is low.

## CHAPTER 5

### STUDIES ON CLONED SEQUENCES FROM A. SATIVUM



## STUDIES ON CLONED SEQUENCES FROM A. SATIVUM

### The isolation and restriction mapping of clones containing B1000-related sequences

Attempts to generate libraries of A. sativum genomic DNA fragments inserted into  $\lambda$  or cosmid vectors were unfortunately unsuccessful. Consequently the plasmid vectors pBR 322 and its derivatives pAT 153 and pBR 328 were employed, although this necessarily resulted in a reduction in the size of genomic fragments that could be successfully cloned.

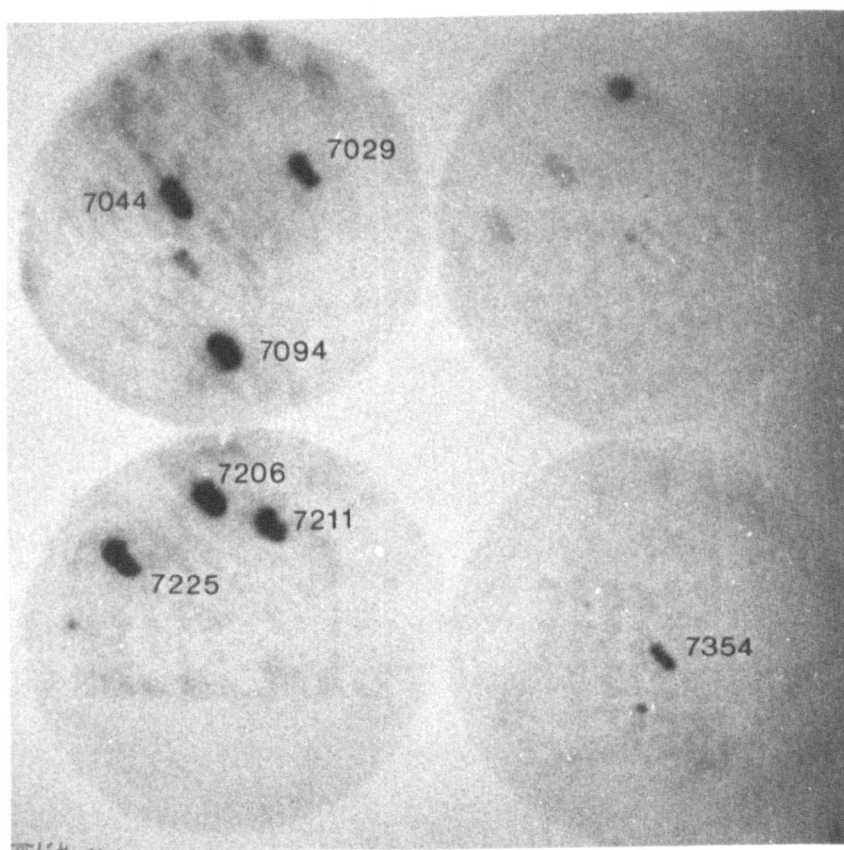
A partial shotgun library of A. sativum Hind III fragments was constructed and the recombinants screened with B1000 probe by the method of Grunstein and Hogness (1975), as described in Materials and Methods.

Figure 12 shows the result of such a screen. Several colonies produced very strong hybridisation signals, and were accompanied by ones of less intense signal. Presumably a weaker signal implies that sequences homologous to the B1000 probe are either shorter or more diverged than those of stronger signal.

All positively hybridising clones were analysed initially by sizing on agarose gels. The size of inserts ranged up to 5.4 kb (data not shown) with

FIGURE 12

Partial Hind III clone library of A. sativum  
DNA probed with  $^{32}\text{P}$  B1000. The numbers of the  
hybridising clones are marked.



a distribution skew towards smaller fragments as might be expected from kinetic considerations inherent in random shotgun cloning methods. The fact that all these plasmids hybridise to the original B1000 probe may suggest that they all share some sequence homology and, depending upon the extent of sequence divergence, might therefore be expected to possess common restriction sites over some of the insert length. This would not necessarily be the expectation if, for example, each of the clones carried a different portion of the B1000 sequence, in which case there might be no homology between the clones. From data presented earlier however it does appear as though the B1000 sequence exists as a contiguous unit at most genomic sites.

Six of the positive clones (pAs 7044, -7029, -7094, -7206, -7225 and -7354) were subjected to further analysis by restriction mapping using the enzymes Bam HI, Eco RI, Hind III and Bgl II. There was no apparent homology between clones in the restriction sites of these four enzymes (see figure 13), perhaps surprising considering their strong hybridisation to a common probe. Nor did the restriction maps of any of the clones seem to bear any relationship to that of the cloned B1000 sequence used as the original probe; all lack a 1050 bp Bam HI fragment.

The ultimate test of sequence homology would be the sequencing of the inserts, although a fine

restriction map ought to reveal any homology. Consequently fine restriction maps were constructed of the B1000 probe and the positive clones using the enzymes Hae III and Alu I, both enzymes having four base-pair recognition sites. This was achieved by partial digestion of  $^{32}\text{P}$  end-labelled fragments as described in Materials and Methods. The restriction maps are presented in figure 13.

Furthermore, to narrow down the particular regions of interest, the restriction fragment(s) responsible for the hybridisation to the B1000 probe was identified by probing Southern blots of a series of different digests of each plasmid with B1000. One example, that of pAs 7044 is shown in figure 15.

The maximum length of the region possessing homology to B1000 probe is presented for each recombinant plasmid in figure 13, denoted by (----).

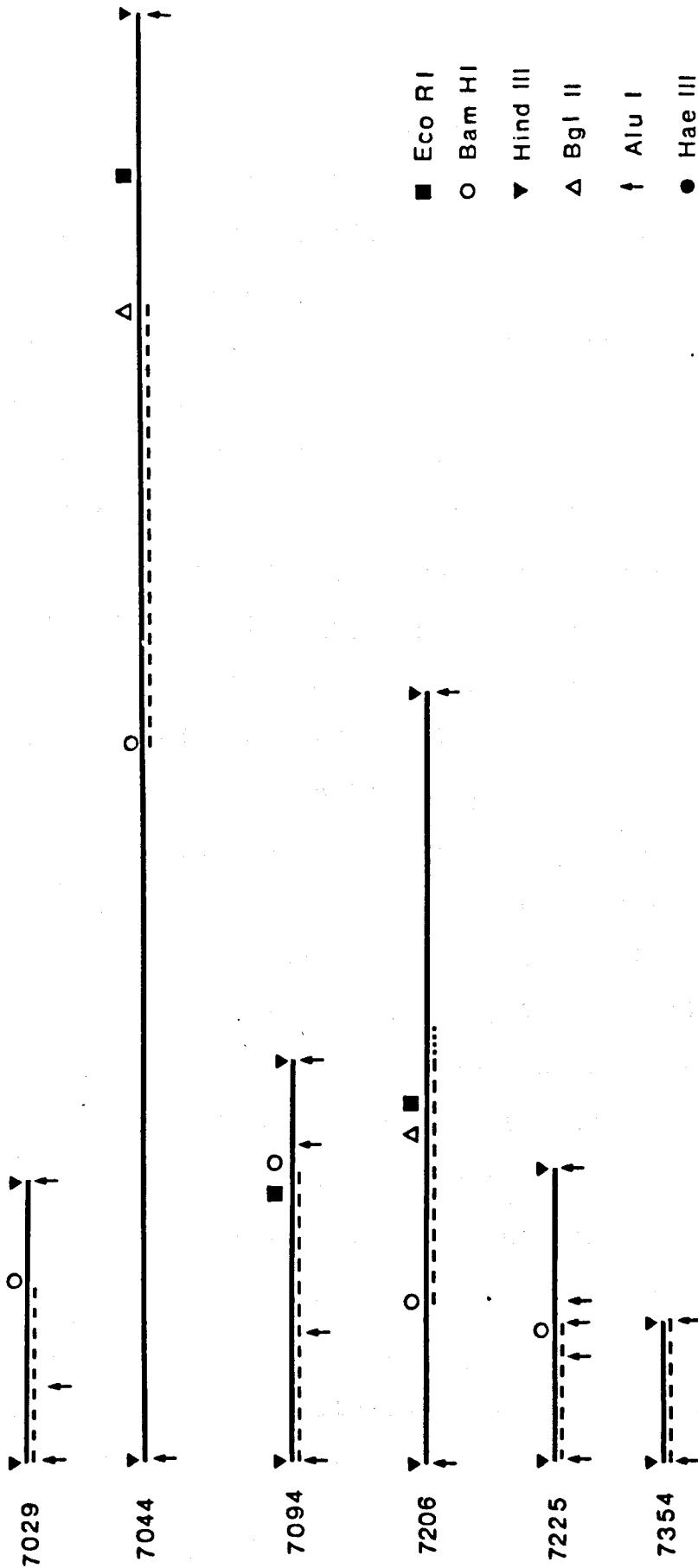
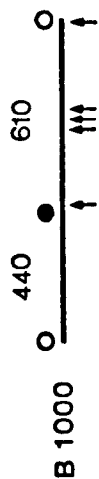
There is clearly not total concordance between the restriction maps of B1000 sequence and the clones, nor between the clones themselves.

There appears to be a paradox. At the level attainable by the probing of Southern blots of digested genomic DNA there appears to be a fairly high degree of homogeneity in B1000 and its flanking sequences, as demonstrated by the rather discrete bands of hybridisation on autoradiographs such as

FIGURE 13

Restriction maps of positively hybridising  
A. sativum Hind III clones.

(The Hae III and Alu I sites internal to pAs 7044  
and pAs 7206 were not determined).



in figure 6(c). However at the somewhat finer level of restriction mapping of clones there appears to be little concordance of restriction sites either within the fragment hybridising to B1000 or within the fragments flanking it.

It is of course possible that in the small sample, organisations of B1000 which are atypical of the common orientations have been cloned. Some of the inserts of the clones are however of the same length as bands highlighted by B1000 in Hind III digests of A. sativum (see figure 23). This suggests, but by no means proves, that each plasmid is a cloned example of a configuration of B1000 that is common in the A. sativum genome. If so the homogeneity observed at one level is in contrast to the heterogeneity observed at another.

These observations can be explained by the existence of microheterogeneity. The insertion, deletion, inversion or amplification of one or many sequences of very short length would be obvious at the level of fine restriction mapping but less so at the higher level of Southern blot analysis. At the latter level small increases or decreases in the length of individual sequences would not be detectable, even if restriction sites were introduced or deleted, unless occurring in a large number of sequences.

If re-amplification of sequences is a common



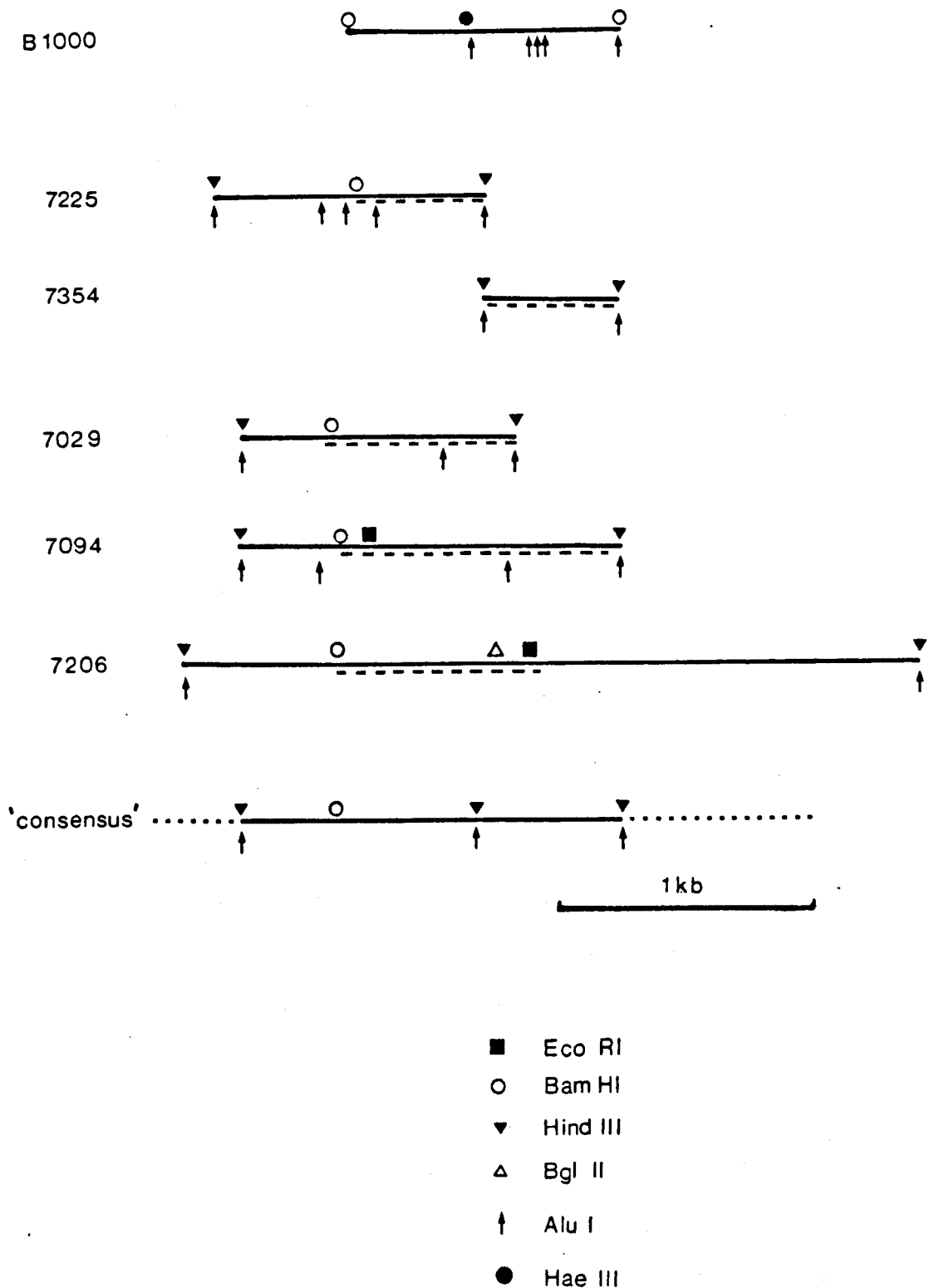
phenomenon in the genome, a variant sequence might be detectable by Southern blot analysis if for example a restriction site was deleted prior to amplification up to the level of detection. Such an occurrence could explain the presence of the many discrete strong bands on autoradiographs such as presented in figures 6(c), 7, 21, 22 and 23.

Sequence rearrangements on a small scale, leading to the phenomenon of microheterogeneity have been observed in members of a number of repetitive sequence families.

Dubnick et al (1983) subjected ten cloned examples of the mouse 1.3 kb Eco RI family to restriction mapping. Although certain regions of the repeats seemed to be highly conserved there was intraspecific sequence divergence. In regions that were sequenced, this divergence was found to be due to single base substitutions, and insertions or deletions of one or a few bps.

Ten clones of the MIF-1 family were sequenced by Brown & Piechaczyk (1983) and found to display sequence variation as a consequence of not only individual base changes but also small insertion sequences and tandem repetitions.

Accepting that there may be a certain degree of microheterogeneity within the sequences involved, it is possible to construct a tentative map of a



**FIGURE 14**

Optimal overlap alignment of the restriction maps of figure 13.

'consensus' element in which B1000 resides.

In figure 14 the restriction maps of B1000 and the B1000-containing recombinant plasmids have been aligned so that there is an optimal degree of restriction site concurrence.

If the concurrence is real the data suggests that there is a repetitive element of which B1000 is a part that, despite being somewhat heterogeneous in sequence at different sites in the genome, retains sufficient aspects of its internal structure to be recognisable by DNA hybridisation and restriction mapping studies. Unfortunately the small size of the clones used in this study do not permit an extension of the consensus element beyond about 1.5 kb with any confidence. Finer mapping of the larger clones might facilitate such an extension as would alignment of fragments of larger length cloned in  $\lambda$  or cosmid vectors.

That B1000 is part of a larger repetitive element that is somewhat heterogeneous in sequence is supported by the data presented in Chapter 4. It also provides an explanation for the patterns of hybridisation produced when digested genomic DNA is probed with B1000 (figures 6(c), 7 and 9). Although the data of figure 10(b) suggests that the many discrete bands on such autoradiographs are not liberated by partial digestion of a larger repeated element they could be liberated from such

an element if a subset of copies represented sequence variants. Such a subset, which may have gained or lost certain restriction sites due to mutational events before their amplification (or due to homogenisation afterwards, see Chapter 8 ) might, upon digestion, liberate different sized bands showing homology to B1000 than do non-variant sequences. Such a situation would explain the different degrees in intensity of many of the bands of hybridisation in figures 6(c), 7 & 9 for example; these being due to differences in copy number of sequence variants as a result of differential amplification.

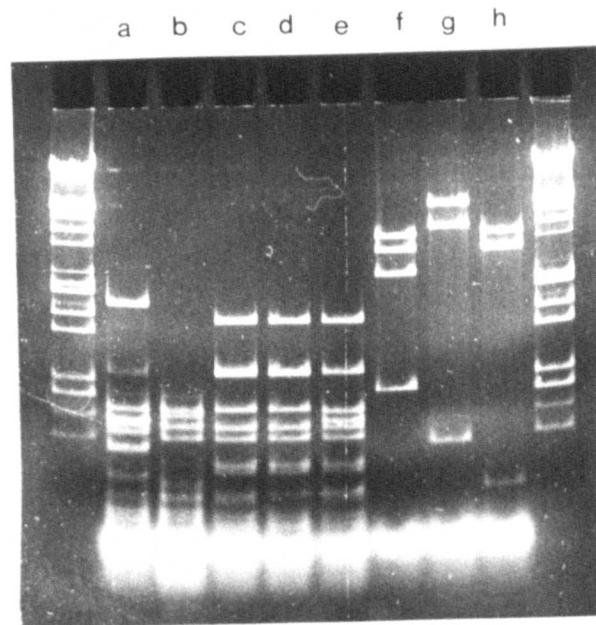
In a simpler situation it might be possible to map common restriction sites external to the B1000 sequence and thus construct a consensus restriction-map of the larger element (see for example Manuelidis and Biro, 1982). Unfortunately due to the apparent complexity of the organisations involved, and the paucity of restriction sites for endonucleases recognising 6 bp (i.e. not too frequent) sites internal to the probe sequence, this cannot be achieved.

#### The sequence environment of B1000-related sequences

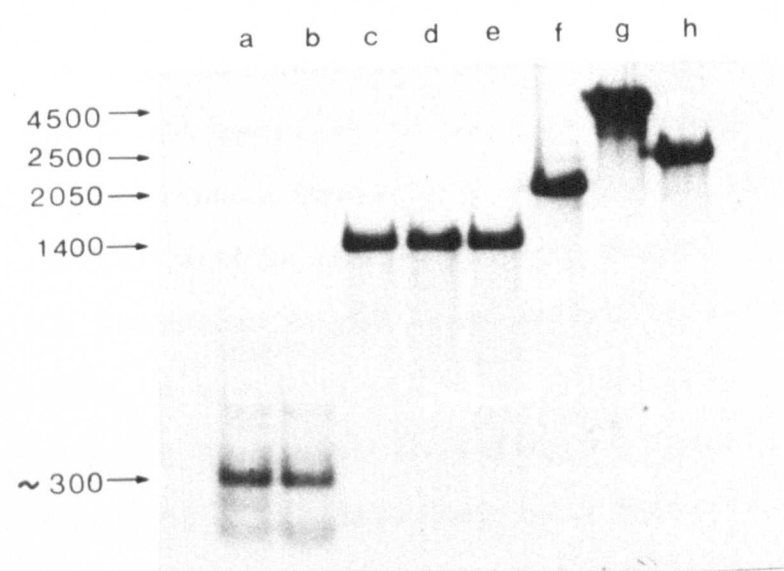
In two of the recombinant plasmids, pAs 7044 and pAs 7206, the sequence responsible for hybridisation to the probe is totally enclosed within

FIGURE 15

- a) pAs 7044 digested with (a) Alu I,  
(b) Hae III/Alu I, (c) Hae III, (d) Bam HI/  
Hae III, (e) Eco RI/Hae III, (f) Bam HI/  
Eco RI, (g) Hind III/Eco RI, (h) Hind III/  
Bam HI.
- b) DNA of figure 15(a) Southern blotted  
and hybridised with  $^{32}\text{P}$  B1000.



a



b

regions of A. sativum DNA that show no homology with, and are presumably unrelated in sequence to, B1000. This is evidenced by lack of hybridisation of B1000 probe to restriction fragments either side of the fragment displaying hybridisation (for example see figure 15.).

This permitted the analysis of sequences flanking B1000-type repeats in the genome of A. sativum. It was also possible to establish whether the flanking sequences in one recombinant plasmid showed any homology to the corresponding flanking sequences in another. Were the B1000 sequence part of a larger repetitive element, within which it was always embedded, then flanking sequences in one recombinant plasmid might be expected to show homology to those in another. Alternatively, were the B1000 sequence to be a relatively short repetitive element found in a variety of sequence environments within the A. sativum genome the flanking sequences might be expected to vary from clone to clone.

In order to pursue this line of investigation, clone pAs 7206 was subdivided into three fragments I, II and III, of length 600, 750 and 1550 bp respectively, by triple digestion with Hind III, Bam HI and Eco RI (see figure 13). The three fragments were fractionated on an agarose gel, electroeluted and nick-translated for use as probes.

Initially each probe, I, II and III, was used to probe Southern blots of various restriction digests of pAs 7206, the plasmid from which they were obtained. Such a 'back-hybridisation' was necessary to determine whether there was any cross-hybridisation between probes I, II and III. There was an absence of hybridisation of any of the three probes to anything other than the restriction fragments predicted from a study of the restriction map of pAs 7206 (data not shown). This suggests that there is no internal repetition within the insert, unless of course it is restricted to being totally within one of the three probe fragments or is of very short length. In both these cases such internal repetition would not be detectable by 'back-hybridisation'.

The above experiment also demonstrates that there is no cross-contamination of fragments by any others. The probes I, II and III can therefore be considered essentially pure.

Each probe was hybridised separately to

- a) The partial library of Hind III generated recombinants from which the six plasmids were originally isolated
- b) Southern blots of restriction digests of A. sativum genomic DNA.

The resulting autoradiograms are presented in figure 16.

Some of these data are presented in tabular



FIGURE 16

Subfragments of pAs 7206 used as probes  
against :-

a) a partial Hind III library of A. sativum  
DNA

b) total A. sativum DNA digested with

(a) Alu I, (b) Hae III and (c) Hind III.

The length of common prominent bands is marked  
in kilobases.

pAs 7206

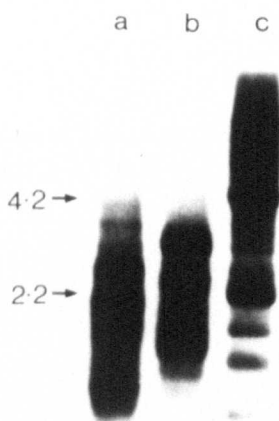
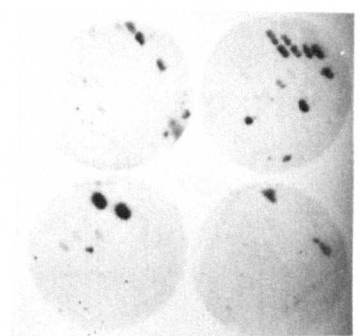
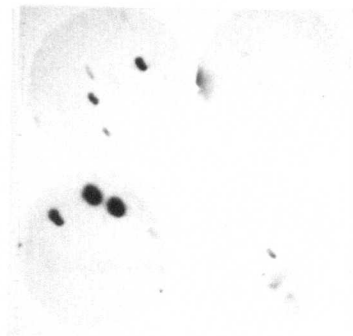
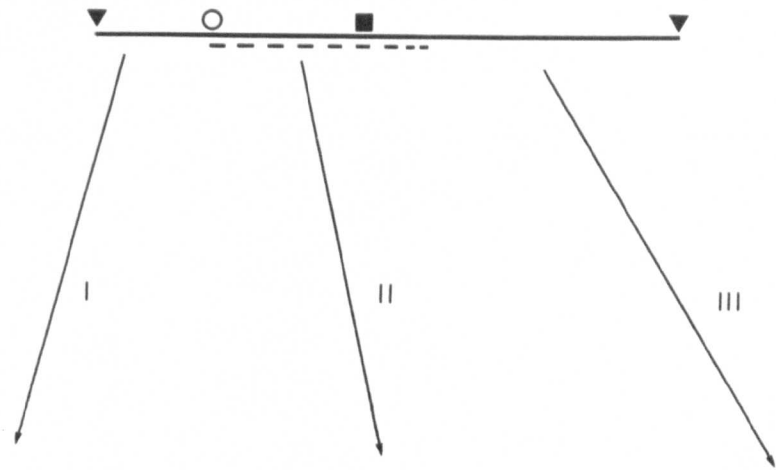


TABLE 2

probe	I			II			III		
	●●●	●●	●	●●●	●●	●	●●●	●●	●
clone	7029	7225	7180	7029	7044		7206	7001	7002
NOs	7206		7001	7068	7094		7211	7018	7044
				7206	7354		7164	7105	7045
	7211		7044	7211				7106	7068
				7225				7109	7142
								7110	7232
								7111	7234
								7112	7246
								7302	
								7327	

TABLE 3

probe	A			B			C		
	●●●	●●	●	●●●	●●	●	●●●	●●	●
clone	7044			7044	7211		7044		
NOs				7180	7206		7180		

●●● denotes very strong hybridisation

●● denotes strong hybridisation

● denotes weak hybridisation

|| denotes the duplication of a clone

form in table 2, where the clones picked out by each probe and the intensity of hybridisation in each case is detailed to ease comparison.

Clearly the probes I, II and III are repetitive. Each hybridised to clones other than pAs 7206 (from which they were isolated) suggesting the presence of homologous sequences at a number of different sites in the genome.

Confirming the presence of a B1000 related sequence in fragment II, probe II hybridised to the same clones as does the original B1000 probe. A number of these clones were also picked out by probes I and/or III (for example pAs 7044, pAs 7029) suggesting that DNA sequences present in fragments I, II and III are found contiguously arranged at other sites in the genome; evidence again for a larger repeating unit.

There were a number of clones which hybridised to only one of the probes. Were the plasmid inserts long this might be indicative of one fragment being present independently of the other two at some sites. Therefore the insert length in a number of these clones was established by restriction mapping. Most were too small to enable the above conclusion to be made, being < 1kb in length (data not shown).

That fragments I, II and III were repetitive

was further suggested by the fact that each hybridised strongly to a large number of fragments, and highlighted many discrete bands, on Southern blots of digested A. sativum genomic DNA (figure 16). The intensity of hybridisation appeared to be similar in each case when specific activity of probe and DNA concentration on each filter were taken into consideration.

The pattern of hybridisation for each of the three probes I, II and III, is however different. Differences would be expected due to differences in the internal sequence of each probe and consequently in those fragments to which each hybridises. This would be particularly noticeable in digests involving enzymes with 4 bp recognition sites such as Hae III and Alu I which liberate smaller fragments than those with 6 bp recognition sites. This is indeed reflected in differences in pattern in the digestion products, particularly those below ~2 kb.

However if fragments I, II and III are often found contiguously arranged in the genome it would be expected that there would be larger bands in common between digestion products probed with each, i.e. hybridisation to fragments liberated by restriction at sites external to the element composed of I, II and III.

Such a pattern is observed. In Hind III digested DNA, all three probes hybridise to major bands of 2.2 kb and 4.2 kb as well as to a number of common minor bands. Similarly all three probes hybridise to fragments of 3.0 kb and 3.1 kb in Hae III digests.

The hybridisation patterns are too complex to enable the construction of a restriction map of any B1000-containing element with any confidence. Nevertheless it can be concluded that sequences flanking a B1000 related sequence in clone pAs 7206 are themselves repetitive, to approximately the same degree, and seem to be found associated with B1000 sequences at other sites in the genome. This suggests the existence of a conserved repetitive element of length at least 4.2 kb (the length of the largest common band visible in hybridisation of each fragment to digested A. sativum genomic DNA ).

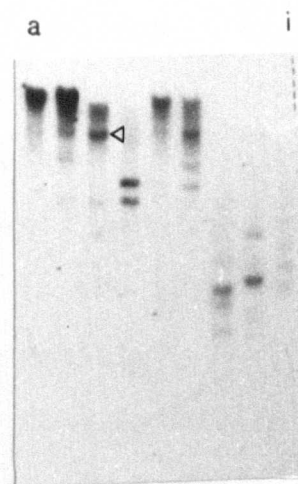
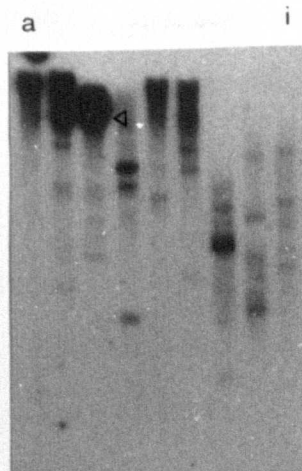
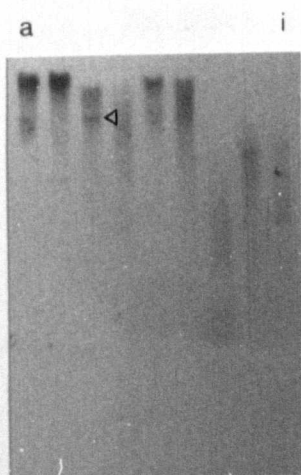
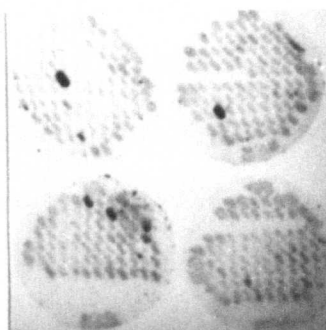
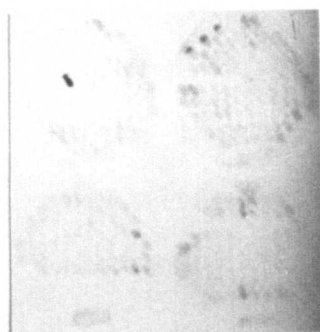
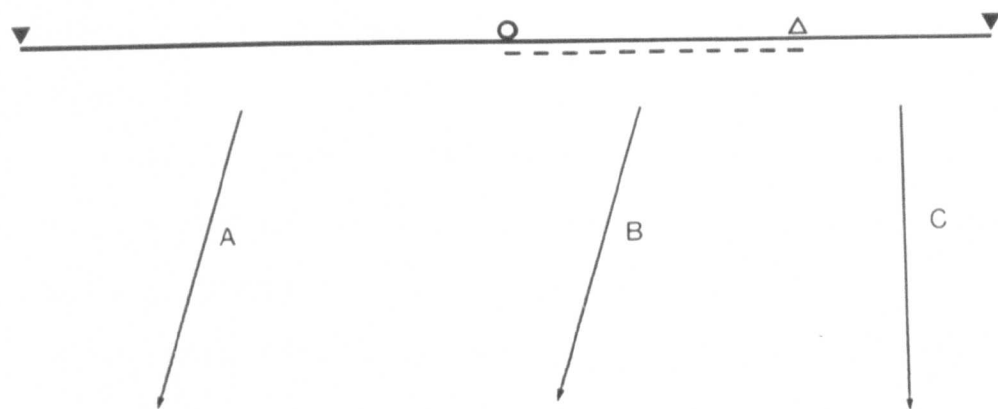
In a similar fashion to the above, the insert of clone pAs 7044 (5.4 kb) was subdivided into three fragments, A, B & C by triple digestion with Hind III, Bam HI and Bgl II as shown in figure 17. Each was nick-translated and used as probes in the same way as probes I, II and III above. The resulting autoradiograms are presented in figure 17 and some of the data is presented in tabular form in table 3.

FIGURE 17

Subfragments of pAs 7044 used as probes  
against :-

- a) a partial Hind III library of  
A. sativum DNA
- b) total A. sativum DNA digested with  
(a) Hind III, (b) Bam HI, (c) Eco RI,  
(d) Bcl I, (e) Bgl II, (f) Hinc II,  
(g) Hae III, (h) Alu I and (i) Hinf I.

pAs 7044





Probes A, B and C hybridise strongly to many fewer clones than do I, II and III, perhaps surprising considering their length. In fact probe A only hybridises strongly to the clone from which it was originally isolated.

Although the sample number of clones is low this might be taken to suggest that sequences making up the bulk of fragment A are of lower repetition frequency than those of I or III for example.

Unexpectedly, fragment B of pAs 7044, the fragment to which the B1000 probe showed homology in the initial screen does not hybridise to the same clones as the corresponding B1000-homologous fragment II in pAs 7206 or as the original B1000 probe; merely to a subset of them. This might suggest that only a part of the B1000 probe sequence is contained within fragment B.

As with pAs 7206, the evidence from the hybridisation patterns produced by probing digested genomic DNA with subfragments of pAs 7044 suggest that such subfragments are found contiguously arranged at many sites in the genome, i.e. there is evidence of a large repetitive element. All three probes from pAs 7044 hybridise to a > 6 kb band in the EcoRI digest for example( 4 ), while many other major bands are common to digests probed with at least two of the probes (for example

the major bands in a Bcl I digest probed with B and C).

The conclusion is that in both the larger clones pAs 7206 and pAs 7044, B1000 sequence (or part of it) is flanked by sequences that are contiguous with it at many other sites in the genome and that restriction sites external to it are conserved to a certain degree. This is highly suggestive of a conserved element in which B1000 sequence resides. Is B1000 always present in the same element? If this were the case then sequences flanking B1000 in one clone ought to be at least partially homologous to sequences flanking it in another if the sequence of the larger element is conserved to a large enough degree.

That there is some homology between the two clones is suggested by the fact that probes I, II and III hybridise to pAs 7044 (albeit weakly in the cases of I and III) and probe B hybridises to pAs 7206 in the filter hybridisations of figures 16 and 17 (see also tables 2 and 3). The homology between the two clones was further investigated by probing Southern blots of various restriction digests of pAs 7044 with fragments I and III of pAs 7206. The results are shown in figure 18.

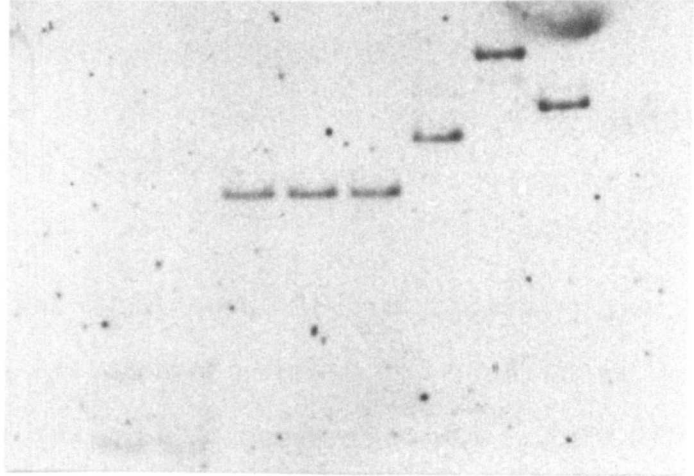
Homology between pAs 7044 DNA and probes I and III is somewhat limited, as evidenced by the long exposure times necessary in order to detect any hybridisation signals; both probes however hybridise to a different subset of bands.

FIGURE 18

pAs 7044 digested with combinations of several restriction endonucleases (see legend to figure 15) and hybridised with

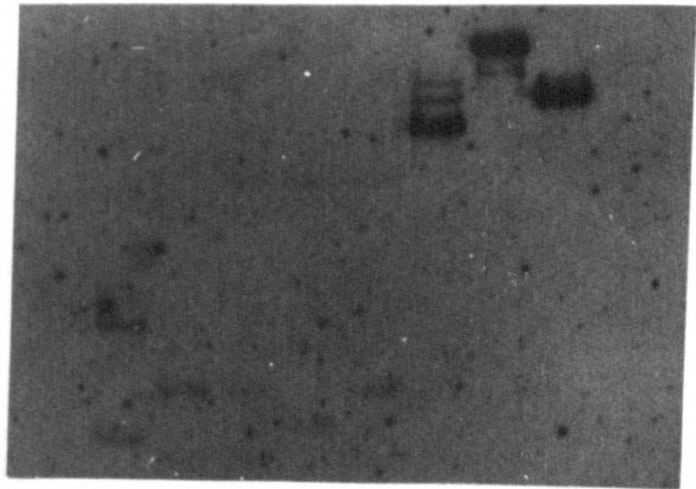
- a) probe I of pAs 7206
- b) probe III of pAs 7206

a b c d e f g h



a

a b c d e f g h



b

Although information obtainable from these data is limited due to the paucity of convenient restriction sites within pAs 7044, it does appear as though homology between the plasmids extends beyond the Bl000-related sequence within each and into at least one of the flanking fragments of pAs 7206.

How far this homology extends and how well conserved are the sequences demonstrating homology cannot be determined from the above data.

The relatively weak cross-hybridisation between the two clones and the apparent difference in the Bl000-homologous sequences within each might be an indication that the two are members of different families of Bl000-containing elements or are simply very diverged in sequence from one another. Finer restriction mapping, further cross-hybridisation studies and even sequencing may be required to determine the relatedness of the two inserts.

## CHAPTER 6

THE CHARACTERISATION OF SEQUENCES  
HOMOLOGOUS TO B1000 IN RELATED ALLIUM SPECIES

THE CHARACTERISATION OF SEQUENCES HOMOLOGOUS  
TO B 1000 IN RELATED ALLIUM SPECIES

(i) The repetitive sequence components of  
related Allium species

A comparison between the repetitive sequence components of related genomes can be made using Cot curve and interspersed curve analysis. Examples of such studies and their limitations were detailed earlier (see Introduction).

A comparison of the physical organisation of DNA in related genomes can also be made by observing the fluorescence patterns of digested genomic DNA fractionated on a gel and stained with ethidium bromide. The value of this approach tends to decrease with increasing C-value, in the sense that large complex genomes may display less discrete bands of fluorescence than comparatively smaller ones. Nevertheless, such an approach has been employed to study gene amplification, enabling a vast array of amplified DNA fragments to be identified in mouse cells for example (Bostock and Tyler-Smith, 1982). The digested genomic DNAs of cereals have many fluorescence bands in common when displayed in the aforementioned way, suggesting that they share many conserved repeated sequences (Bedbrook et al., 1980a); many of these are highly repeated tandemly arranged DNA

sequences (Jones and Flavell, 1982).

The fluorescence patterns of Bam HI digested and fractionated total DNA from several Allium species is presented in figure 19. There are clearly differences in the number and the distribution of fluorescence bands between many of the species, particularly between A. vineale and A. sativum for example which are regarded as being rather closely related. Presumably a difference in fluorescence pattern reflects differences in the repetitive sequence component between two genomes. Of particular interest therefore is the presence in A. vineale and A. cepa of a ladder of fragments whose lengths are integral multiples of 400 bp. Such an arrangement is typical of a satellite DNA component and from the fluorescence patterns alone such a component appears to be present in some species yet absent from others (such as A. sativum). Data relating to a satellite component of some Allium genomes is presented later (Chapter 7 ).

Those bands common to several species (the 'ladder' for example in A. vineale and A. cepa) presumably, but do not necessarily, comprise related sequences, suggesting that at least a part of the repetitive sequence component of one species is present in another. The relatedness of sequences cannot of course be established from such data alone, although a similarity of fluorescence patterns is suggestive of shared sequences..

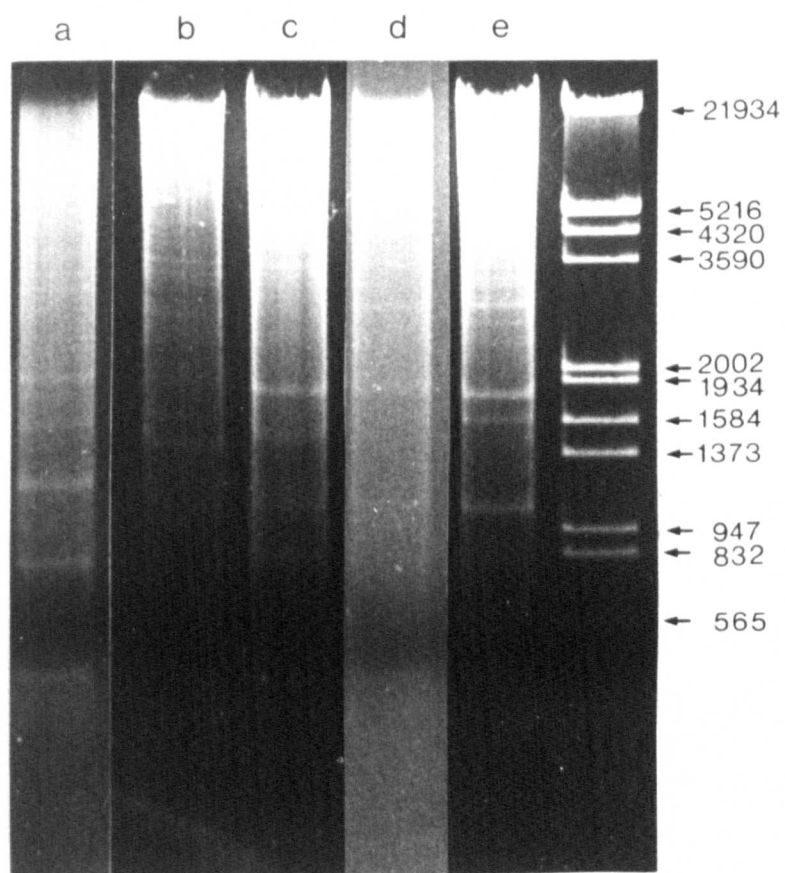


FIGURE 19

Total DNA from several species of Allium digested with Bam HI and fractionated on an agarose gel.

- a) A. vineale
- b) A. cepa
- c) A. babingtonii
- d) A. porrum
- e) A. sativum

The lengths of bands of a  $\lambda$ Eco RI/Hind III marker are given in base pairs.



Nevertheless, it is clear that the repetitive sequence components of even rather closely related species can be rather different as judged by differences in restriction patterns of total DNA.

(11) The prevalence of B 1000 in other *Allium* species

The proportion of each genome represented by DNA fragments residing in the 1050 bp band (the source of the B 1000 probe from *A. sativum*) can be estimated in each case by densitometry. Such an approach is however of little use in studying the prevalence or physical organisation of a specific repetitive sequence family such as B 1000 in the genomes of different species, simply because other sequences in each genome homologous to B 1000 cannot be determined by such means.

An indication of the prevalence of B 1000-like sequences in genomes of other *Allium* species is given by the degree of hybridisation of B 1000 probe to "spot blots" of DNA samples. This method (Cunningham, 1983) offers the advantage of accurate quantification of DNA by its fluorescence on ethidium-bromide plates prior to blotting, as described in Materials and Methods.

Because hybridisation to nitrocellulose bound DNA is performed under fairly stringent conditions (65°C in 4 x SSC) the degree of hybridisation will be determined not only by the quantity of homologous sequences but also by the degree of mismatching between probe and

genomic sequence. It has been estimated that every 1% mismatching reduces the  $T_m$  of a duplex by about  $1^\circ\text{C}$ ., although the magnitude depends upon the base composition (Bonner et al., 1973).

Genomic DNA from 19 different species/varieties of Allium (see table 1 ) were spotted onto ethidium bromide plates, as was a dilution series of  $\lambda$  DNA standards. DNA was visualised under UV illumination permitting a quantification of each sample to be made (figure 20(a)). After partial depurination, DNA was denatured, neutralised and hybridised with  $^{32}\text{P}$  B 1000 as described in Materials and Methods. The resulting autoradiograph is presented in figure 20 (b). After the variation in quantity of DNA in each sample is taken into consideration, there is clearly considerable variation in the degree of hybridisation of B 1000 to the DNA of each species. Strongest hybridisation is to A. sativum, the source of the probe, and to the apparently very closely related species of A. babingtonii, A. porrum, A. vineale, and A. sphaerocephalon.

Hybridisation is however very weak to the apparently more distantly related species A. triquetrum ( $2n=18$ ), A. ursinum ( $2n=14$ ), A. moly ( $2n=14$ ) and A. neapolitanum ( $2n=14$ ) all of which differ from A. sativum in their basic chromosome number. Also comparatively weak is hybridisation to species apparently closely related to A. cepa

## FIGURE 20

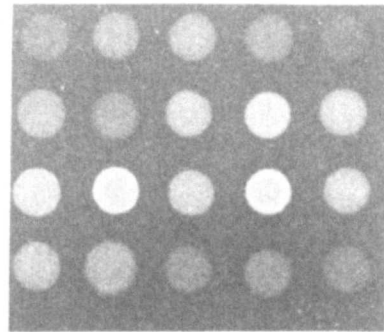
Spot blot hybridisation to DNA from Allium species

- a) Ethidium bromide stained DNA samples
- b) Autoradiograph of the above following transfer to nitrocellulose filter and hybridisation with  $^{32}\text{P}$  B1000.

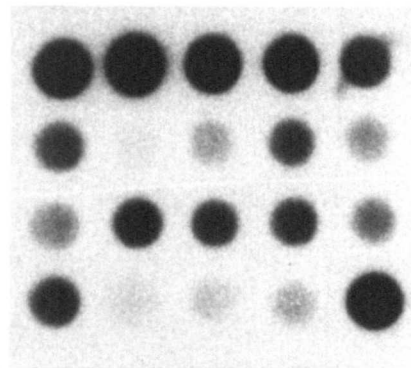
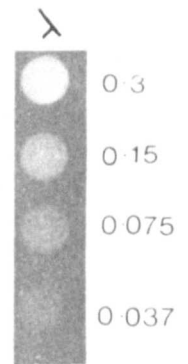
The DNA samples are from:

- |                              |                                     |
|------------------------------|-------------------------------------|
| 1. <u>A. sativum</u>         | 11. <u>A. fistulosum</u>            |
| 2. <u>A. porrum</u>          | 12. <u>A. pekinense</u>             |
| 3. <u>A. sphaerocephalon</u> | 13. <u>A. schoenoprasum</u>         |
| 4. <u>A. babingtonii</u>     | 14. <u>A. tuberosum</u>             |
| 5. <u>A. vineale</u>         | 15. <u>A. neapolitanum</u>          |
| 6. <u>A. ostrowskianum</u>   | 16. <u>A. stipitatum</u>            |
| 7. <u>A. triquetrum</u>      | 17. <u>A. moly</u>                  |
| 8. <u>A. ursinum</u>         | 18. <u>A. cepa</u> var spring onion |
| 9. <u>A. cepa</u>            |                                     |
| 10. <u>A. hierochuntium</u>  | 19. <u>A. cepa</u> var tree onion   |
|                              | 20. <u>A. sativum</u>               |

A  $\lambda$ -DNA dilution series is also presented. The amount of DNA is marked in  $\mu\text{g}$ .



a



b

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20

such as A. hierochuntium and the cepa varieties 'spring-onion' and 'tree-onion'.

Accurate quantification of the degree of prevalence of B1000 in these species is rendered difficult without an idea of the degree of sequence mismatching involved; intensity does however seem to be in accordance with phylogenetic relationships as presented in figure 3.

(iii) The organisation of B1000 in other Allium species

An indication of the physical organisation of B1000-like sequences in the genomes of other species was obtained by probing Southern blots of digested and fractionated DNA from each species with  $^{32}\text{P}$  B1000.

As with spot blots, the amount of hybridisation to each DNA is determined by the result of a combination of the effects of copy number and sequence divergence, and as a consequence neither value can be accurately determined without an estimate of the other. However the pattern of hybridisation is less likely to be affected by these two parameters and can yield information on the probe's physical organisation in genomes other than that of A. sativum.

The DNA of 10 species was further investigated in this way. Approximately 2 $\mu\text{g}$  of each DNA was digested separately with Bam HI, Eco RI and Hind III, fractionated on a 1% agarose gel, Southern blotted onto nitrocellulose filter and probed with B1000.

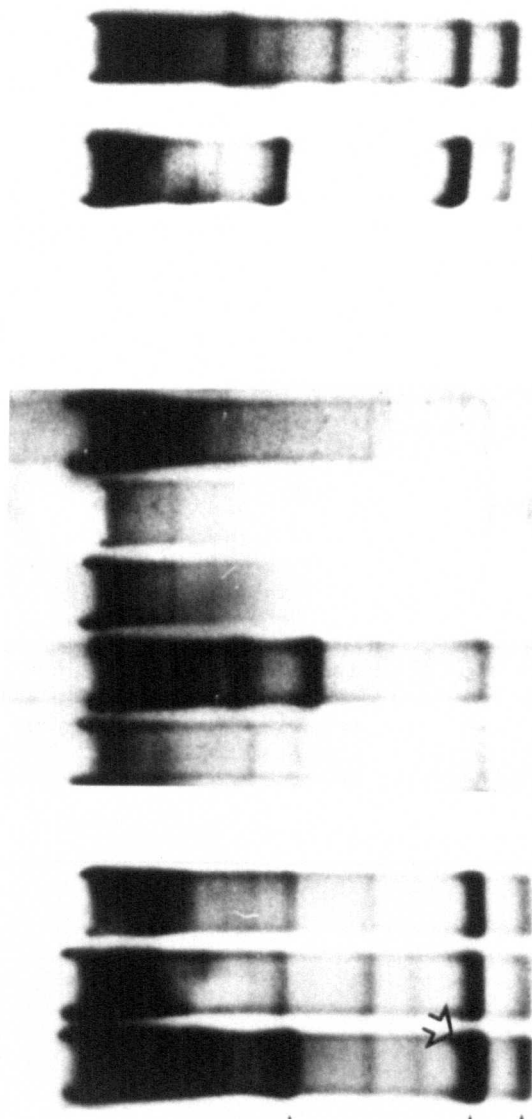
FIGURE 21

Bam HI digested DNA from (a) and (i) A. sativum, (b) A. babingtonii, (c) A. porrum, (d) A. triquetrum, (e) A. schoenoprasum, (f) A. cepa, (g) A. cepa (spring onion variety), (h) A. tuberosum, (j) A. vineale, and (k) A. cepa (tree onion variety) hybridised to  $^{32}\text{P}$  B1000.

Lanes (a) to (h) are from the same original gel, lanes (i), (j) and (k) from a different gel. Lanes (d) to (h) were exposed for five times the period of the other lanes in order to show the very weak hybridisation patterns. The length of some of the prominent bands is marked in base pairs, the position of the original 1050 bp probe sequence is marked with an arrow.



a b c d e f g h i j k



2900

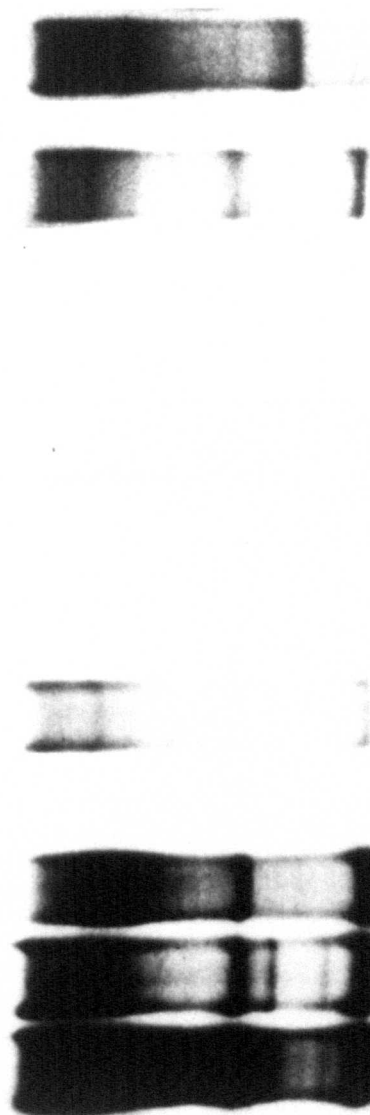
1050

840

FIGURE 22

Eco RI digested DNA from several Allium species  
(see figure 21) hybridised with  $^{32}\text{P}$  B1000.

a b c d e f g h i j k

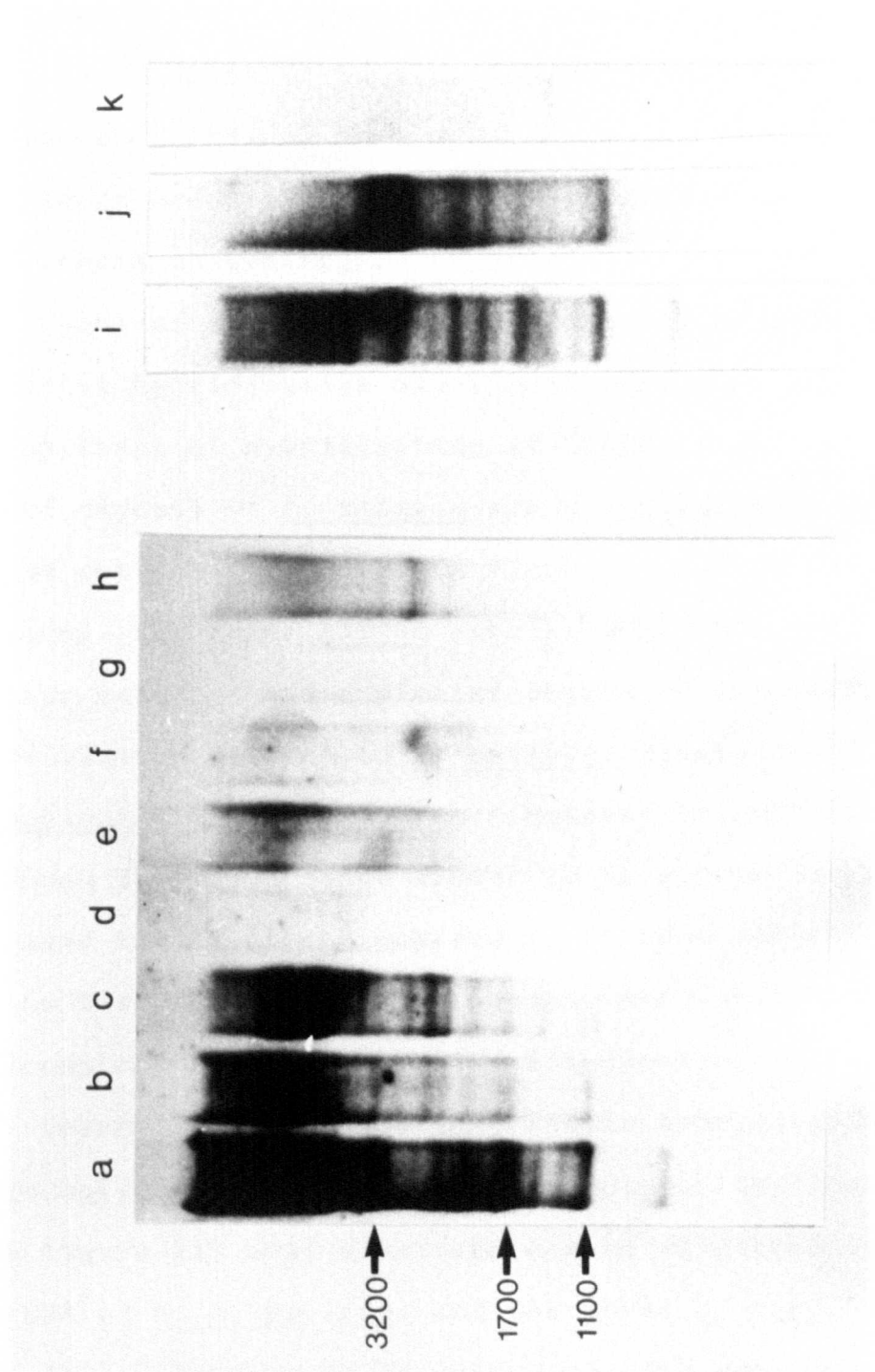


2750

1450

FIGURE 23

Hind III digested DNA from several Allium  
species (see figure 21) hybridised with  $^{32}\text{P}$  B1000.



The resulting autoradiographs are shown in figures 21, 22 and 23. The lanes (d), (e), (f), (g) and (h) in figure 21 were exposed for 5 times as long as all other lanes in order to show the very weak hybridisation patterns.

There are very obviously considerable quantitative differences in the degree of hybridisation to DNA of each species; differences which reflect those in the spot-blot hybridisation experiment earlier (figure 20). The patterns of hybridisation of B1000 to Bam HI and Eco RI digests of A. sativum are essentially identical to the patterns presented in figure 6 (c) and discussed earlier. Two other Allium species' DNAs show very similar patterns and a similar degree of intensity of hybridisation as that to A. sativum, namely A. babingtonii and A. porrum. However in both cases although the major bands appear to be approximately the same length, there appears to be some variation in the distribution of minor bands, most particularly noticeable perhaps in the Hind III digests of figure 23.

Several species show very little hybridisation. It is only on rather over-exposed autoradiographs (see figure 21) that a pattern can be discerned in the DNA of A. triquetrum, and the three A. cepa varieties. However it is interesting to note that in all these examples a band of length 1050 bp, i.e. the length of the original B1000 probe, is observable in Bam HI digests, as are weak bands

corresponding to other major bands observable in A. sativum digests. Thus it appears that although the degree of hybridisation of the probe to the DNA of these species is greatly reduced for whatever reason, the pattern seems to have been conserved to some extent.

It is the third situation that may be of most significance. Although in both A. vineale and A. schoenoprasum bands of 1050 bp are highlighted by B1000 in Bam HI digests, many of the other major bands in the three digests are of different length to those found in corresponding digests of A. sativum. This observation is particularly pronounced in A. vineale. The degree of hybridisation is very similar to that to A. sativum, but in all three digests most of the bands of hybridisation are of different length to those in A. sativum.

It might be useful to classify each of the 10 DNAs investigated in this way into one of three general groups:

1. Genomes to which B1000 hybridises to approximately the same degree as to A. sativum DNA producing substantially the same pattern of hybridisation.
2. Genomes to which B1000 hybridises to a very much reduced degree compared to with A. sativum but which nevertheless show some similarity in their patterns of hybridisation.

3. Genomes to which B1000 hybridises and reveals a substantially different pattern of hybridisation compared to A. sativum.

Group 1 includes A. sativum, A. babingtonii and A. porrum, Group 2 A. triquetrum, A. cepa, A. cepa var Spring onion, A. cepa var tree onion and A. tuberosum, and Group 3 A. vineale and A. schoenoprasum.

Although all species fall reasonably easily into these three groups it must be noted that there are minor differences between even the most similar patterns of hybridisation. Furthermore because hybridisation to DNA of the species assigned to Group 2 is only detectable on the more over-exposed autoradiographs, patterns of hybridisation are not always clearly visible. In those instances where a pattern is visible, differences from that of A. sativum are apparent in all cases. The significance of these observations will be discussed later.

Hybridisation of B1000 to discrete bands of fragments in the fractionated digestion products of species other than A. sativum (see figures 21, 22 and 23) suggest that in these species too, B1000-related sequences are frequently found adjacent to a limited number of other repetitive sequences. This conserved, non-random arrangement of restriction sites external to B1000-like sequences in a number of different genomes might suggest that B1000 exists



in such genomes within a larger conserved repetitive element. Unfortunately the patterns of hybridisation are too complex to enable the construction of a consensus sequence of any such element with any confidence.

(iv) An estimate of Bl000 sequence divergence between species

Any sequence mismatching will affect the degree of hybridisation in such experiments as those outlined above and could conceivably affect the pattern of hybridisation. This latter effect would depend on whether Bl000 sequences in certain configurations in the genome are any more or less homogeneous than those in other configurations.

To determine the effect of sequence mismatching upon results such as those presented in figures 21, 22 and 23, the effect of increased and decreased stringency upon the degree of hybridisation between Bl000 and heterologous DNA was investigated.

DNA from A. sativum, A. porrum (an apparently closely related species) and A. cepa (apparently more distantly related) was digested separately with Bam HI and Hind III, fractionated on a 1% agarose gel, transferred to nitrocellulose paper and probed with Bl000 at 45°C., 65°C. and 75°C., representing low, medium and high stringency respectively. All nine pairs of digests (3 species, 3 temperatures) were hybridised separately to minimise any competition

effects between different DNAs. All lanes had the same quantity of DNA and the same amount of probe applied to them and the resulting autoradiographs are presented in figure 24.

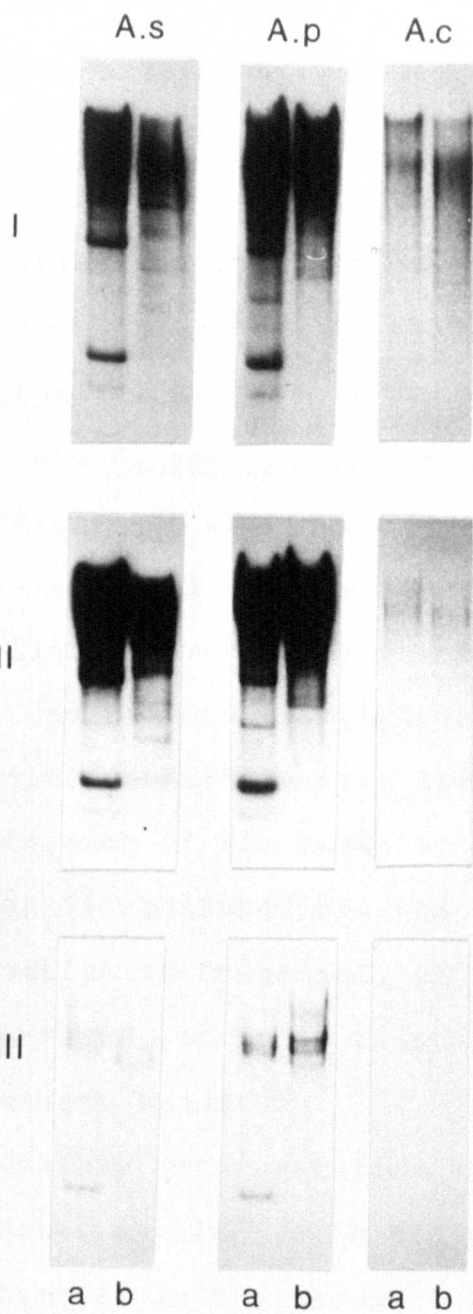
As might be expected, most hybridisation to A. sativum DNA occurs at 65°C., which is closest to the optimum temperature for reassociation under these conditions (usually regarded as  $T_m - 25^\circ\text{C}.$ ). At 65°C. there is also strong hybridisation to A. porrum DNA, the pattern of which closely resembles that of A. sativum DNA; a result in accordance with the data from spot blots (figure 20(b) and Southern blots of digests (figures 21, 22 and 23). Also in accordance with these data is the observation that at 65°C., hybridisation to A. cepa DNA is considerably reduced compared with that to A. sativum.

However at the reduced hybridisation stringency of 45°C., slightly less hybridisation to A. sativum and A. porrum DNA occurs while that to A. cepa is increased compared to that at 65°C. This is presumably because imperfectly matched duplexes are able to form between B1000 and A. cepa sequences at this reduced stringency, duplexes which were prevented from forming stably at 65°C. This immediately suggests that a quantity of B1000-related sequence exists in the A. cepa genome that is not detectable under standard reassociation conditions (4xSSC, 65°C.).

The fact that hybridisation to A. cepa at

FIGURE 24

Southern blots of DNA from A. sativum (A.s),  
A. porrum (A.p) and A. cepa (A.c) digested with  
(a) Bam HI and (b) Hind III hybridised with  
 $^{32}\text{P}$  B1000 at I 45°C., II 65°C. and III 75°C.  
representing low, medium and high stringency.



reduced stringency is in the form of a smear rather than of discrete bands implies a large degree of heterogeneity in Bl000-like sequences in this genome. A large degree of heterogeneity would also result in increased mismatching and the reduced degree of hybridisation observed at 65°C.

At the increased stringency of 75°C., virtually all hybridisation to A. cepa DNA is abolished, again suggesting that Bl000 related sequences in the genome of this species are somewhat more diverged from the probe sequence than those in the genomes of A. sativum and A. porrum. Indeed at 75°C., hybridisation to A. sativum and A. porrum, although considerably reduced compared to that occurring at 65°C., is still detectable, suggesting the presence of sequences which are little diverged from the Bl000 probe sequence. The pattern of hybridisation at 75°C. to these two species differs from the pattern at 65°C. in so far as much of the background high molecular weight smear is abolished and the relative proportion of hybridisation to fragments of length 1050 bp is greatly increased, as it is to some other higher molecular weight bands.

The possible interpretation that sequences sharing greater homology with Bl000 reside within specific elements in the genome will be discussed later.

In conclusion, this experiment suggests that while the B1000-related sequences in the genome of A. porrum have been conserved to approximately the same degree as those in A. sativum during the evolution of the two species, any homologous sequences in the genome of A. cepa have undergone a considerable degree of divergence. The same conclusions can be drawn from a study of the thermal stability of duplexes formed between B1000 and homologous sequences in the genomes of four Allium species.

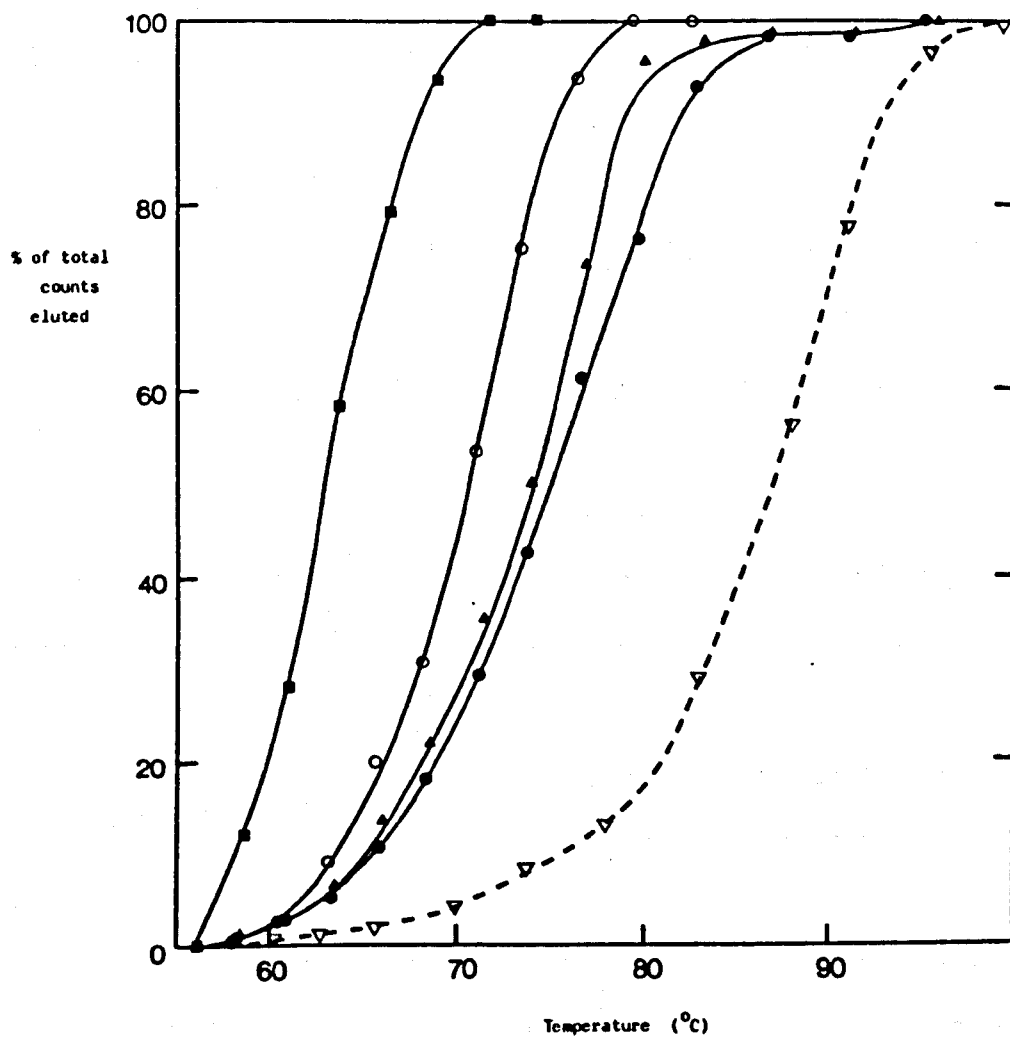
$^{32}\text{P}$  B1000 and an excess of sheared, unlabelled genomic driver DNA from either A. sativum, A. porrum, A. cepa or A. vineale were denatured and reassociated in solution to a Cot of 21.5 at 60°C. Reassociated duplex molecules were immobilised on HAP in a water-jacketted column and TCA precipitable counts (representing denatured single stranded B1000 DNA) were monitored during incremental rises in temperature. The results for each of the species are presented in figure 25, the counts at each temperature expressed as a proportion of the total number of counts eluted in each experiment.

As expected, duplexes formed between B1000 and homologous sequences in the genome of A. sativum were the most stable, having a  $T_m$  of 74°C. Duplexes involving A. porrum DNA also showed a high degree of

## FIGURE 25

Thermal stability of heteroduplex molecules formed between  $^{32}\text{P}$  labelled B1000 probe and genomic DNA from A. sativum (● - ●), A. vineale (○ - ○), A. porrum (▲ - ▲) and A. cepa (■ - ■).

A melting curve of native A. sativum DNA is also presented ( - - - ). Denaturation was monitored by absorbance of DNA at OD 260 with incremental rises in temperature. Values are expressed as a percentage of the total increase in OD.





thermal stability, having a  $T_m$  only  $1^{\circ}\text{C}$ . lower than A. sativum, suggesting an approximate 1% mismatch. This is entirely in accordance with the data presented above (figure 24), as is the fact that B1000:A. cepa DNA hybrid molecules appear to be very unstable with a  $T_m$  of  $63^{\circ}\text{C}$ . being some  $11^{\circ}\text{C}$ . lower than B1000:A. sativum. This represents a mismatching of some 11%.

Interestingly B1000:A. vineale duplexes, although apparently less well matched than those formed between B1000 and A. sativum or A. porrum DNA, nevertheless show only a  $3.5^{\circ}\text{C}$ . difference in  $T_m$ , representing 3.5% mismatching. Therefore there appears to be considerable homology between B1000 and A. vineale DNA despite the fact that some considerable differences in the physical organisation of B1000 in A. vineale (compared to within A. sativum and A. porrum) is suggested by such data as that in figures 21, 22 and 23.

The implications of this will be discussed later.

A melting curve of native A. sativum DNA is also presented in figure 25. Sonicated total DNA was loaded onto HAP at  $60^{\circ}\text{C}$ . in 0.12M PB and single stranded DNA eluted at each incremental rise in temperature. The OD 260 of each fraction was measured and expressed as a fraction of the total for the experiment. A  $T_m$  of  $86^{\circ}\text{C}$ . suggests a G + C base pair content of approximately 40%, (Marmur and Doty, 1959; Lewin, 1980), similar to the value

reported by Kirk et al. (1970) using chemical methods. The  $\Delta T_m$  of over  $10^{\circ}\text{C}$ . between native A. sativum DNA and B1000:A.sativum duplexes implies a certain degree of heterogeneity of those sequences able to hybridise with B1000 probe under standard conditions. The heterogeneity is however less than that observed between B1000 and DNA from species other than A. sativum.

The sequence divergence of B1000-like sequences in the A.sativum genome appears to be average for moderately repeated sequences in that genome. A thermal elution curve was constructed for A.sativum DNA that had been reassociated at  $60^{\circ}\text{C}$  to  $Cot\ 10$ . Although showing a slightly broader melting profile than the curves in figure 25 , such reassociated DNA had a  $T_m$  of approximately  $77^{\circ}\text{C}$ . (data not shown), a value not dissimilar from the  $T_m$  of B1000:A.sativum duplexes.

A degree of sequence divergence suggested by thermal elution studies is in accordance with the data from hybridisation at different criteria (figure 24) as well as the restriction map heterogeneity reported earlier (see figure 13).

## CHAPTER 7

### A SATELLITE-DNA COMPONENT OF A. VINEALE

## A SATELLITE-DNA COMPONENT OF A. VINEALE

### Identification, isolation and characterisation of a satellite component in A. vineale

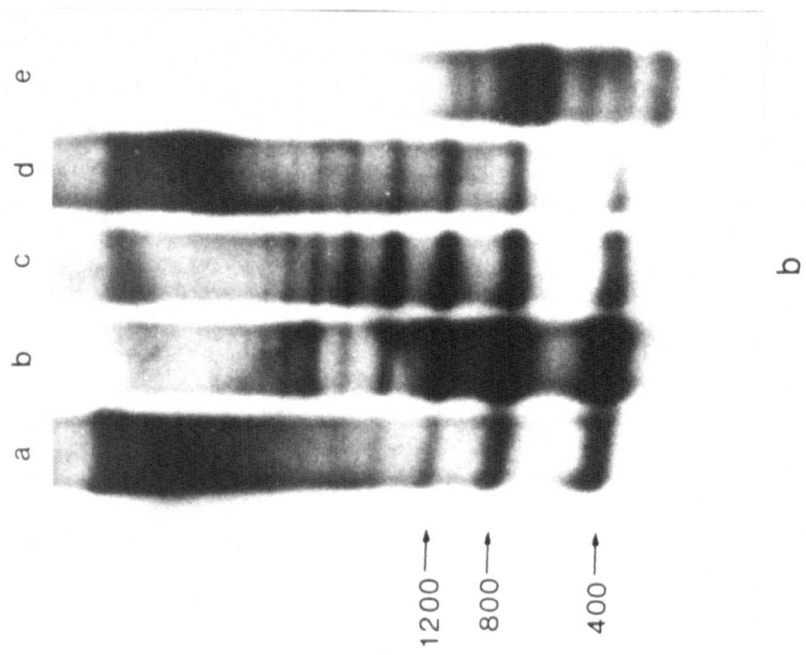
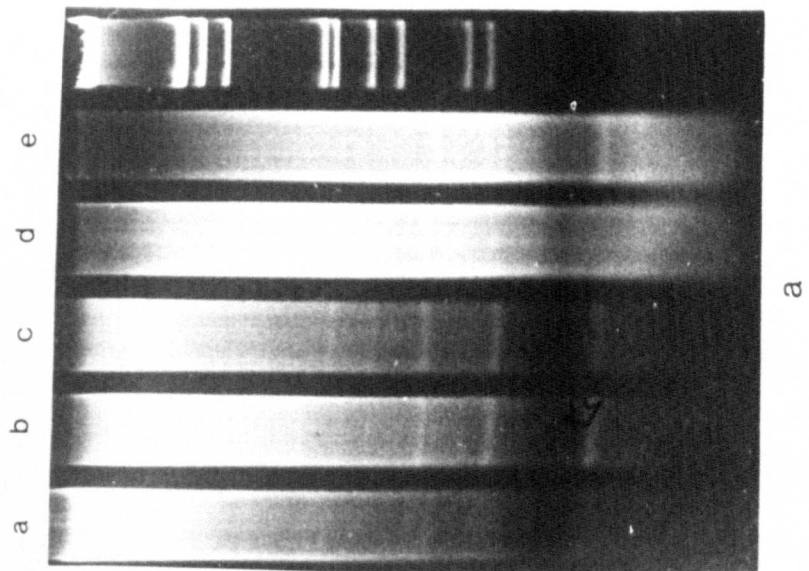
The existence of a satellite-type tandem array of repetitive DNA sequences in the genome of A. vineale was initially suggested by the presence of a ladder of fragments in ethidium bromide stained, fractionated digestion products of genomic DNA. Several restriction endonucleases liberated such a pattern, among them Hind III, Eco RI and Bst I (the latter being an isoschizomer of Bam HI). Examples are shown in figure 27(a), the lengths of the fragments forming the ladder being integral multiples of 400 bp. Presumably the 400 bp fragments represent the monomer, the 800 bp fragment the dimer and so on. Fragments of 2400 bp representing 6-mers are clearly visible in ethidium bromide stained gels.

In order to confirm the arrangement of such sequences in the genome of A. vineale, genomic DNA was digested with Eco RI, fractionated on an agarose gel and the 400 bp monomer band excised and nick-translated. This was then used to probe A. vineale DNA which had been digested with a number of different enzymes and transferred to nitrocellulose filter. The resulting autoradiograph is shown in figure 27(b). The characteristic satellite ladder

FIGURE 27

- a) A. vineale DNA digested with (a) Bst I, (b) Eco RI, (c) Hind III, (d) Hae III, (e) Alu I, and fractionated on an agarose gel. The marker is an Eco RI/Hind III digest of  $\lambda$  (see figure 19).
- b) Autoradiograph of the above following transfer of DNA to nitrocellulose filter and hybridisation with  $^{32}\text{P}$  nick-translated 400 bp band DNA (arrowed in (a)).

Bands representing the monomer, dimer and trimer fragments are sized in base pairs.

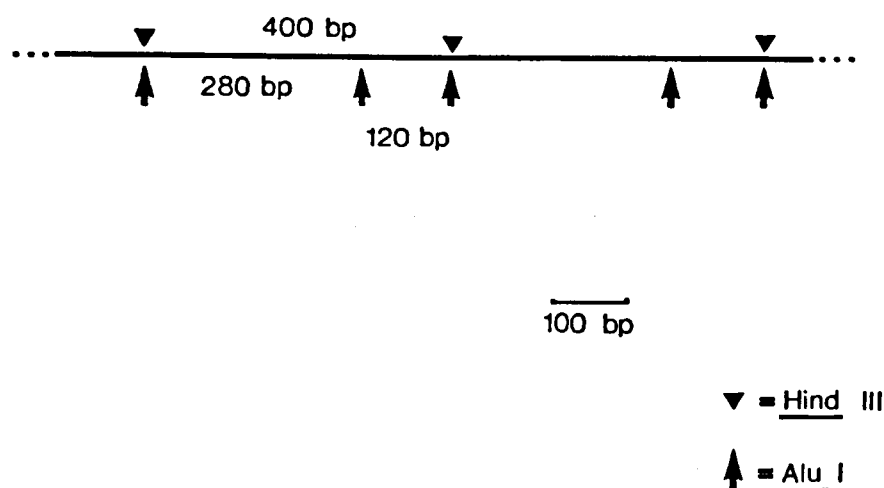


is indeed produced in four of the digests, confirming that fragments that are multiples of 400 bp in length are indeed homologous to the 400 bp monomer. Furthermore the fact that hybridisation of the monomer is only to fragments that are integral multiples of 400 bp suggests that there is only one recognition site for each of the endonucleases Hind III, Eco RI, Bst I and Hae III. The ladder pattern is presumably created as a result of the loss of recognition sites for each enzyme by mutation; bands representing 9-mers are clearly visible in figure 27(b).

The hybridisation pattern resulting from hybridisation of the monomer unit to Alu I digested A. vineale DNA is however different, there being a more complex ladder which contains fragments of length smaller than the 400 bp monomer, (see figure 27(b)). This immediately suggests that there is more than one Alu I recognition site within the 'consensus' unit sequence; a fact which permits the partial restriction map of figure 28 to be constructed. The alignment of Alu I sites in relation to Hind III sites is possible because Alu I recognises the central four base-pairs of Hind III recognition sites and thus will always cut at sites recognised by Hind III. All the bands in the Alu I digested A. vineale DNA track of figure 27(b) can be accounted for as being partial

digestion products of DNA conforming to the restriction map presented in figure 28. The sites for the other enzymes cannot however be assigned from this data alone, simply because all produce the same sized fragments.

FIGURE 28



To obtain purer probe, cloned monomer unit DNA was sought. A partial library of A. vineale Bst I fragments in pBR328 was screened by the method of Grunstein and Hogness (1975) with  $^{32}\text{P}$  labelled 400 bp monomer band. Of several positive clones one, pAv 1101 was used as a source of the 400 bp monomer for further work, it having been shown to contain a 400 bp Bst I fragment homologous to the monomer probe by hybridisation to Southern blots of restricted recombinant plasmid (data not shown).



The distribution of the 400 bp satellite in related Allium species

The distribution of sequences homologous to the A. vineale satellite component among other Allium species was determined by hybridisation of  $^{32}\text{P}$  labelled pAv 1101 to a spot blot of DNA from 19 different species/varieties (the same spot-blot as used in figure 20). Strongest hybridisation was to A. vineale DNA, the source of the satellite probe. Weaker but distinct hybridisation occurs to DNA from A. babingtonii, A. sphaerocephalon, A. porrum and A. schoenoprasum (data not shown). Whether this was due to differences in amount or in sequence divergence cannot be established from this data.

It was surprising that there was little or no hybridisation to DNA from A. cepa. As alluded to earlier, the genome of A. cepa contains a satellite component consisting of a 400 bp array; this DNA is isolatable as a satellite band on a CsCl density gradient but appears to be less simple in its organisation as judged by the ladders of fragments produced upon hybridising cloned monomer to digested A. cepa DNA, (A. James, S.R. Barnes, pers.comm.). Although the satellite components of A. vineale and A. cepa have the same repeat length they do not appear to be closely related.

The organisation of homologous sequences in related Allium species

The DNA of several species was digested with Bst I, fractionated on an agarose gel, transferred to nitrocellulose and hybridised with  $^{32}\text{P}$  400 bp cloned monomer. The results of hybridisation to the DNA of 9 different Allium species is shown in figure 29.

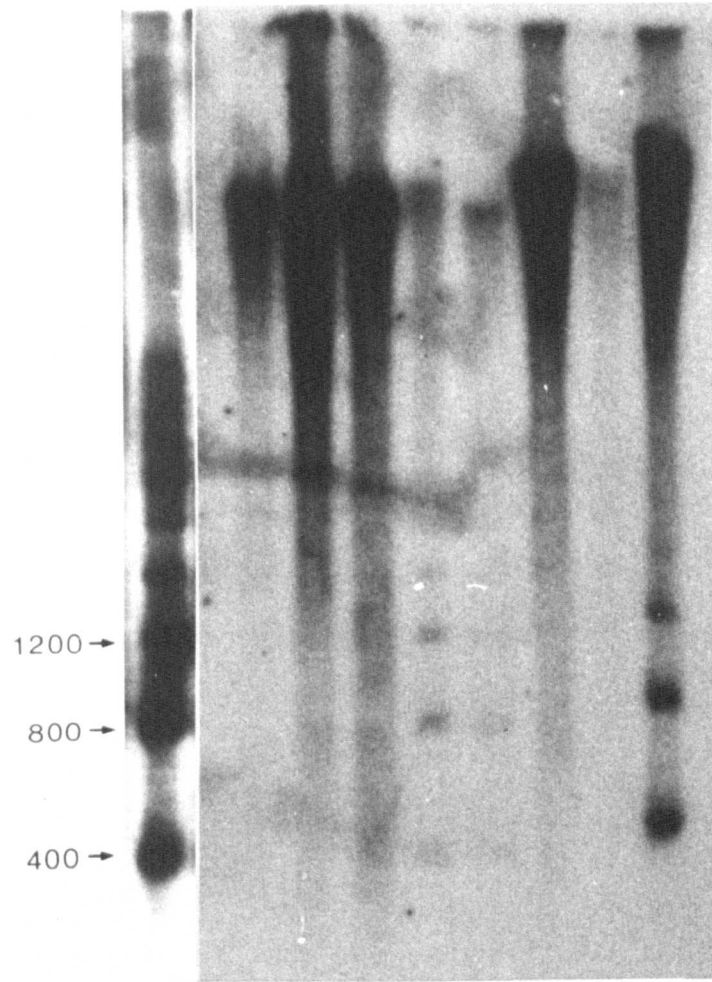
There is hybridisation to all 9 species; the degree of hybridisation however varies considerably. Interestingly, although there is reasonably strong hybridisation to the DNA from A. sativum, A. babingtonii, A. porrum and A. schoenoprasum it is largely confined to high molecular weight DNA. Conversely hybridisation to A. cepa and A. cepa var, spring onion is largely to a ladder of fragments, while that to A. triquetrum is to both a ladder and high molecular weight DNA.

Where ladders exist in DNA from species other than A. vineale, hybridising fragments are of lengths representing multiples of 400 bp. Thus any sequences sharing homology with A. vineale satellite in these species appear to have the same basic organisation as in the genome of A. vineale. These results strongly suggest that at least some of the satellite components of A. vineale and A. cepa are related, sharing some homology in

FIGURE 29

Total DNA from (a) A. vineale, (b) A. sativum, (c) A. porrum, (d) A. babingtonii, (e) A. cepa, (f) A. cepa var. spring-onion, (g) A.schoenoprasum, (h) A. tuberosum and (i) A. triquetrum, digested with Bst I (= Bam HI), fractionated, Southern blotted and hybridised with  $^{32}\text{P}$  400 bp A. vineale DNA. Similar results were obtained using  $^{32}\text{P}$  pAv1101 as a probe. The length of the bands representing the monomer, dimer and trimer fragments are marked in base pairs.

a b c d e f g h i



both sequence and organisation; the degree of homology is however rather limited.

Hybridisation of the A. vineale 400 bp monomer probe to DNA fragments of high molecular weight from species such as A. sativum after digestion with Bam HI could be interpreted as suggesting either a different organisation of homologous sequences in the latter (i.e. non-tandem) or an absence of Bam HI restriction sites in any tandem arrays present.

The evolutionary history of the various Allium satellite components is not easy to establish from the rather limited data presented above. There is clearly however a degree of between-species heterogeneity and within-species homogeneity as evidenced by data such as that of figures 27 and 29. Models relating to the generation and maintenance of relatively homogenous tandem arrays are discussed later, as is a possible explanation for the patterns of hybridisation obtained in the studies described above.

### Overview

Data relating to the organisation of two repetitive-sequence DNA families in the genomes of species of the genus Allium has been presented.

Members of the B1000 sequence family are organised in a complex fashion in A. sativum, usually found enclosed within a larger yet somewhat heterogeneous element which is interspersed with other unrelated repetitive sequences. The sequence family is also represented in other Allium species to varying degrees; in some, differences in the pattern of hybridisation on Southern blots of restricted total DNA suggest differential amplification of sequences containing B 1000 during evolution.

The organisation and distribution of a second family - the 400 bp 'satellite' family - is different. Isolated from A. vineale, it is present in a different subset of species than B 1000, being absent or in very low copy in A. sativum. Members of this family are organised in tandem arrays, liberating ladders of fragments which are integral multiples of the basic unit repeat (400 bp) upon digestion with a variety of restriction endonucleases.

The physical organisation of many families of repetitive DNA sequences in eukaryotic genomes has been described in the Introduction. For the purposes of the discussion such sequences will be divided into two classes, tandemly repetitive sequences, and interspersed repetitive sequences.

Each of these is further subdivided:

(I) Tandem

- i) Non transcribed 'satellite'
- ii) Transcribed repetitive genes

(II) Interspersed

- i) SINES ) excluding those sequences having  
          ) characteristics of transposable
- ii) LINES ) elements
- iii) Transposable elements

By no means are these categories intended to be definitive subdivisions of repetitive sequences; they are simply based on the physical organisation and physical characteristics of a sequence, and do not necessarily reflect for example differences in biological function.

In subdividing repetitive sequence elements it is not intended that one class be considered in isolation from the others; some genomic processes for example might apply to all. On the other hand, welcome though a unifying theory would be, it seems unlikely that all repetitive sequences can be considered to be subject, in the genome, to all the same processes and pressures.

The subdivision of repetitive DNA sequences is then an attempt to create order from the situation in which, as Bouchard (1982) states ... "the bulk of eukaryotic moderately repetitive elements can at present only be said to be united in their apparant lack of resemblance to transposable elements ..."

The elucidation of the nature and organisation of repetitive DNA sequences in a wide range of eukaryotic genomes raises a number of questions:

- 1) Does the present day organisation suggest any mechanisms for the generation and maintenance of such patterns and of their constituent elements?
- 2) Does a study of the evolutionary history and present day organisation of repetitive DNA sequences suggest a function for such sequences?
- 3) Does the accumulation of repetitive sequences into the organisational patterns observed have any effect upon the functioning of the genome?
- 4) Can the evolutionary behaviour of repetitive DNA sequences be incorporated into the evolutionary theory of speciation?

There follows a review of evidence relating to these questions. Where this evidence is relevant to the data presented for the B1000 and 400bp-satellite families of Allium species (Chapters 4, 5, 6 and 7), discussion is presented. Those sections to which the Allium data does not seem directly relevant ought perhaps to be regarded as signposts to areas where further investigation might prove profitable.



## CHAPTER 8

### THE EVOLUTION OF REPETITIVE DNA SEQUENCE COMPONENTS

## THE EVOLUTION OF REPETITIVE DNA SEQUENCE COMPONENTS

### Possible mechanisms for the generation of organisational patterns

From hybridisation data alone it is possible to conclude that in Allium genomes the physical organisation of the B1000 family and the 400 bp -satellite family differ considerably yet are occasionally present in the same genome. As described in Chapter 7 the satellite sequences appear to be clustered in tandem arrays. The localisation of these clusters is not known and while it is possible that clusters of arrays are themselves dispersed throughout the length of all chromosomes, a very localised heterochromatic distribution might be expected if such sequences are typical of other eukaryotic satellite components.

Members of the B1000 family do not however appear to be clustered in the genome. There is no evidence of tandem arrays of such sequences either in genomic blot experiments or in cloned genomic fragments. The evidence that they are interspersed with unrelated sequences and widely dispersed throughout the genome is further strengthened by the results of in situ hybridisation experiments using  $^3\text{H}$  B1000 to probe the A. sativum chromosome complement. An example is presented in figure 30 (data of G. Jamieson). Hybridisation of the probe occurs at sites along both

FIGURE 30

In situ hybridisation of  $^3\text{H}$  B1000 to metaphase chromosomes of A. sativum (data of G. Jamieson).



10 um

arms of all chromosomes, a situation very atypical of satellite sequences for example. It has not yet been established whether there is any reproducible pattern of silver grains over the complement, an investigation hindered by the difficulties involved in accurately Karyotyping A.sativum chromosomes. Nevertheless sequences hybridising to B1000 appear to be distributed throughout the genome.

How might these different organisational patterns have evolved?

1) Tandem arrays

A study of the present day organisation of satellite DNA in a number of species has enabled the evolutionary history of many families to be established (see for example Biro et al., 1975; Roizes and Pages, 1982).

Although amplification of short sequences appears to be responsible for satellite clusters, the mechanisms involved are not clearly established. Two processes have been proposed, saltatory replication and unequal crossing-over.

Britten and Kohne (1968) first postulated the process of saltatory replication, which describes the sudden replication of a sequence to produce a cluster containing a large number of tandemly arranged copies. The model has since been modified by Walker (1971) and Southern (1975a) in order to

explain the phenomenon of long-range periodicity common to many complex arrays. The modified model, which is presented in figure 31 , postulates that arrays are produced by a series of amplification events - rather than by a single event - with mutation occurring in some copies of the array prior to their re-amplification. The basic unit length of amplified sequences may of course be different at each step.

The actual mechanism of amplification is not clear; it may involve rolling circle replication of excised elements (a process well established in rDNA amplification during meiosis in Xenopus for example, Hourcade et al., 1973) or perhaps aberrant replication of the DNA template. It is in respect of the latter that the phenomenon of gene amplification might be of relevance.

A mechanism by which cells meet the demand for increased quantities of specific gene products, either during development or when subjected to selection pressure in culture, gene amplification appears to be rapid, occurring in a single cell division (reviewed by Schimke, 1982). In at least some cases the mechanism appears to involve unscheduled DNA replication at a single position followed by multiple recombination events which

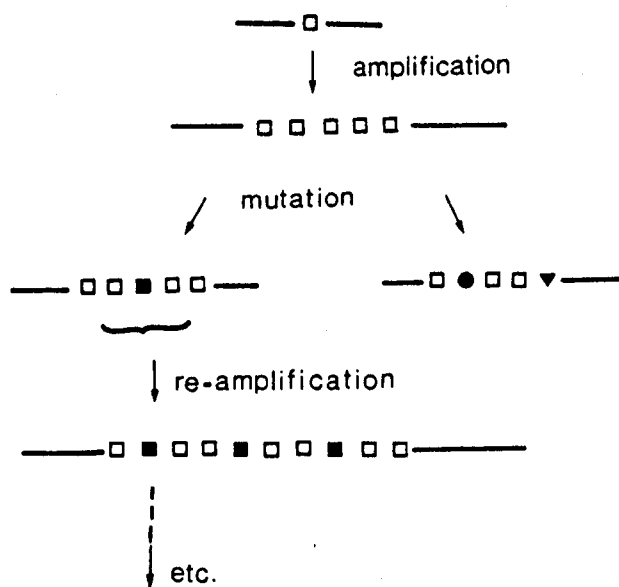


FIGURE 31

Model for the evolution of satellite DNA by saltatory replication.

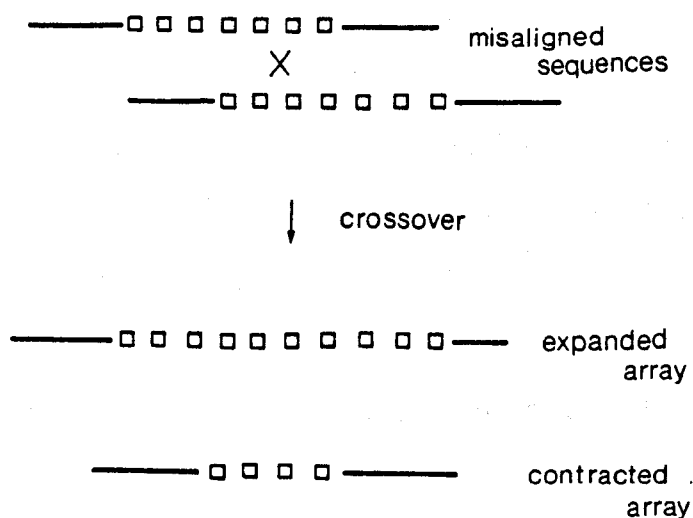


FIGURE 32

Model for the evolution of satellite DNA by unequal crossing over.

resolve the molecules (Roberts et al., 1983). The unit subject to amplification may range in length up to 2000 kb, the resulting array being either unstable, existing as an extrachromosomal double minute chromosome (which has no centromere and is subject to non-disjunction at mitosis) or stable, integrating into the chromosome. If selection for the amplified genes is removed, such arrays are notoriously unstable.

Whether this phenomenon has any bearing upon the generation of satellite arrays is not clear. It does however demonstrate that under certain conditions large scale amplification of genomic sequences can and does occur. Incidentally, even if satellite arrays were to be generated in this way, it would not necessarily imply selection for their sequences. Satellite DNA distribution is unusual in that it is almost invariably heterochromatic. It is possible that the generation of tandem arrays at other positions in chromosomes does occur but that such arrays, unless subject to strong positive selection, are very unstable. In heterochromatic regions however, amplified sequences may not be subjected to the same selection pressures as elsewhere, perhaps being protected by the local DNA configuration (heterochromatic DNA is usually in a condensed state) or simply being tolerated if in a



non-essential region; in this region tandem arrays might be less unstable than elsewhere on a chromosome.

An alternative model for satellite array generation has been postulated which requires frequent unequal crossing over events to occur between repeating units on sister chromatids which have become aligned out of register (Smith, 1976). Computer simulation suggests that stochastic unequal crossing over will eventually lead - as a result of crossover fixation - to a tandem array consisting entirely of a single basic sequence. This model is presented in figure 32. Despite there being difficulties in aspects of both models (see Lewin, 1980) many present day patterns of genome organisation have been explained using schemes which involve the generation of tandem arrays. In the evolution of heterochromatic repetitive sequence elements in cereal species, amplification by unequal crossing-over or rapid saltatory mechanisms is presumed to have been followed by the divergence and dispersal of elements of small size. A cyclic scheme of molecular events for repetitive sequence evolution is proposed (Flavell, 1980) which involves re-amplification of an already amplified and dispersed sequence. If the unit which is reamplified includes, in addition to the original unit of amplification,

unrelated flanking sequences, then the generation of complex or compound repetitive elements can easily be envisaged, (see Flavell, 1980).

There is no reason to suppose that the reamplified complex element will not be dispersed and further reamplified in a different configuration.

The 400 bp satellite-DNA family of A.vineale, reported here, appears to be comparatively simple in its organisation as judged by the ladders of fragments produced by digestion of total DNA. Only fragments representing integral multiples of the 400 bp basic repeat-unit appear when Southern blots of restricted total DNA are probed with either genomic or cloned satellite monomer sequence (see figure 27(b)). No other higher order periodicity is evident in autoradiographs of such blots, although such a pattern might be liberated by restriction endonucleases other than those employed here. A more complex situation is evidently present in the related satellite array in A.cepa. In addition to fragments representing integral multiples of the basic 400 bp unit, there appears to be a minor component consisting of fragments of non-integral length (A. James, S. R. Barnes, pers.comm.)

The relative simplicity of the A. vineale satellite offers no clues as to the mechanism of its generation. The apparent lack of long range

periodicities would seem to suggest relatively few amplification events. The non-integral minor bands in the related family in A. cepa might therefore be evidence for further amplification events in this species subsequent to the evolutionary divergence of A. vineale and A. cepa.

Whether any amplification events proceeded by a process of saltatory replication or unequal exchange cannot be established from the data available. The latter, though, has been implicated as being responsible for copy number variation in the rRNA genes of A. cepa. A 2.5 fold intraspecific variation in the number of rRNA genes in A. cepa has been reported by Maggini et al. (1978) who detected a much smaller variation in A. sativum. Such variation has been detected within a number of plant species (Flavell and Smith, 1974; Cullis and Davis, 1975), and within Drosophila in which the increase or decrease of rRNA genes in response to selection (in the bobbed strain for example, Tartof, 1975) is thought to be due to unequal crossing-over events. Interestingly the difference in the magnitude of variation between A. cepa and A. sativum might be related to the fact that meiotic crossing over does not give rise to polymorphisms in the seed-sterile vegetatively propagating A. sativum. That this fact might have relevance to the organisation of satellite arrays in Allium species is discussed later (page 124 ).

### Interspersed sequences

Neither saltatory replication nor unequal crossing-over can alone adequately explain the mechanism of evolution of interspersed repeated sequence elements in eukaryotic genomes.

Nevertheless, some interspersed sequences may have their origin in tandem arrays. Orphans, a class of dispersed solitary elements derived from tandem multigene families such as histone and ribosomal genes, have been detected in species as diverse as sea urchin, Drosophila and yeast (Childs et al., 1981). While the mechanism is not clear, the presence of a transposable element in a cloned sea urchin orphan (Childs et al., op cit.) might suggest that the orphan was generated as a result of the illegitimate recombination of a transposable-element which had inserted into a tandem array of genes. There is no apparent reason why orphan generation should be restricted to tandemly repetitive gene families; all tandem arrays may be prone to it. Their conclusion that ... "there may no longer be a clear distinction between tandem and dispersed multigene families ..." might also be applicable to non-genic sequences.

In the case of orphans, the parent tandem array is still very much in evidence; indeed the copy number of orphans per genome appears to be limited (Childs et al., op cit.). If a similar mechanism

is to be postulated for the generation of dispersed elements such as MIF-1 and kpn-I, the absence of any parent tandem array must be explained. The data presented earlier provides no evidence for members of the Bl000 family in Allium species being organised in tandem arrays. If such sequences have their origins in tandem arrays then sufficient time must have elapsed since their generation for the entire array to have been dispersed. Tandem arrays of genes (such as histones and rRNA genes) may be subject to more stringent selection for maintenance of configuration than tandem arrays of non-transcribed non-genic sequences (a subject discussed later). It is not inconceivable therefore that the processes to which any tandem arrays are subjected in the genome have a greater effect and act more rapidly upon non-genic sequences.

Indeed, Anderson et al. (1981) in studying cloned members of three different sea urchin repeated sequence families concluded that the organisational status of a particular family may reflect its age. Amplification of a relatively low copy sequence into a tandem array of homogeneous repeats is, they suggest, followed in time by the dispersal of family members and their eventual decay to heterogeneity as a result of the accumulation of mutations.

On a smaller scale, tandem gene duplication

followed by the dispersal and divergence of sequences is thought to be responsible for the nature and organisation of multicopy gene families such as the globins (Jeffreys, 1982) and sea urchin actin genes (Scheller et al., 1981).

Transposition of rather large segments of DNA mobilised by a pair of terminal repetitive elements of the foldback (FB) family has been reported (Paro et al., 1983) and may offer clues as to one mechanism for the dispersal of duplicated or amplified DNA sequences.

Interspersion of repetitive sequences might occur in the absence of transposition mediated events. Roberts et al. (1983) have determined the organisation of those molecules resolved by multiple recombinational events following gene amplification within chromosomes. Their resulting model of gene amplification requires that recombination between irregularly spaced repetitive sequences produces an array of units which are heterogeneous in both size and sequence content. If a region consisting of repetitive sequences were to be subjected to this process the result might be clusters of repeated elements which, due to the multiple recombination events, were related but scrambled in each cluster. Although the mechanism requires that multiple regions of homology must be present within the amplified DNA, it

may be responsible to some degree for members of different families of repetitive sequences (often of heterogeneous length) being interspersed with each other, a phenomenon particularly prevalent in plant genomes as described earlier.

A rather different evolutionary history, one not involving an initial amplification event, has been proposed for some of the copia-like elements. These resemble the proviruses of vertebrate retroviruses in many respects, being of similar length, having direct repeats at each end, and causing a duplication of the target site upon insertion for example (Finnegan et al., 1982). Retroviruses possess the capacity to generate RNA transcripts of themselves, which may become enclosed within virus particles and form infectious agents. Upon infection of a cell, double-stranded DNA copies of the RNA are generated; these are able to circularise and integrate stably into the host genome.

It has been suggested that transposable elements such as those of Drosophila may have evolved from such retroviruses. If so they were presumably not originally part of the ancestral genome but have infected at some stage and become co-adapted to become a heritable part of the present genome in a number of species. The mechanism of transposition of some eukaryotic transposable elements may still resemble that of retroviruses as suggested by the discovery of virus-like-particles containing copia RNA

(Shiba and Saigo, 1983) and of circular DNA containing copia sequences (Flavell and Ish-Horowicz, 1981).

Whether any of the above-mentioned processes are involved in the generation and distribution of many of the large numbers of moderately repetitive interspersed elements in eukaryotic genomes has not been established. The easiest way of explaining the distribution of such sequences would be to credit them with the ability to transpose following their generation. As described earlier some sequences are indeed capable of transposition and it has even been suggested that the majority of middle-repetitive DNA sequences in eukaryotes are, or at one time were, capable of transposition (see for example Doolittle, 1982).

There is indeed evidence for some moderately repetitive DNA sequence families having been in the past transposable but having since lost the ability. The striking similarity mentioned earlier between the copia-like sequences and retroviral sequences also extends to sequences which show no signs of being transposable. Brown (1983) for example reports that sequences homologous to LTR's of retroviruses are found within MIF-1 elements, such sequences including those required for RNA transcription. This might suggest that members of the MIF-1 family were once capable of transposition, possibly via an RNA intermediate (like retroviruses), but that the



majority of members have since accumulated mutations and can no longer support the transcription of RNA necessary for their mobility. There is no reason why this should apply only to MIF-1 family members. The recent discovery of homology between the MIF-1 elements of mouse and the Kpn I elements of primates (Singer et al., 1983) might implicate such a mechanism of generation and dispersal for the major primate LINE family. Although the relationship between the two families is distant, as indicated by their sequence divergence, both appear to have their origin in a common progenitor.

Interestingly retroviral-like sequences have also been found in the repeating unit of a tandemly arranged sequence, one of the bovine satellite DNA's (Streeck, 1982).

Further evidence for transposition of sequences within a eukaryotic genome involving reverse transcription of an RNA intermediate comes from pseudogenes, untranslatable 'decayed' copies of functional genes. (Denison et al, 1981; Chen et al., 1982). Many pseudogenes have lost the intervening sequences of the parent gene and therefore resemble processed RNA products. The integration of reverse transcribed copies of processed mRNA, followed by a relatively rapid accumulation of mutations, could explain the presence, and nature, of many dispersed sequences in eukaryotic genomes (Sharp, 1983).

Replicative transposition of a DNA sequence ensures the self-preservation of a sequence and requires no function to explain its origin or maintenance (other than the function of self-maintenance). The recently introduced concepts of 'selfish' and 'ignorant' DNA are useful in the context of understanding genome organisation.

### Selfish and Ignorant DNA

Selfish DNA (Orgel and Crick, 1980; Doolittle and Sapienza, 1980) is a term which has been used to describe sequences which are present in the genome not as a result of their conferring 'fitness' upon the organism carrying them but simply as a result of their properties of self-maintenance and self-proliferation within a genome. Sequences capable of transposition fit into this category rather neatly. Elements such as the copia-like sequences, Iyl and maybe Alu, all of which have been described earlier, appear to be autonomous in their ability to transpose; their existence is sequence dependent in the sense that certain genes and DNA sequences are required to perform transposition, and are thus presumably maintained by selection. Theoretical studies on the accumulation of selfish DNA have been presented by Ohta (1983) for example.

Some sequences appear however to proliferate

differentially in a sequence independent manner, an example perhaps being the satellite components of genomes discussed earlier. Such sequences have been regarded, like selfish DNA, as conferring no selective advantage upon the individual possessing them. In addition however they have been assumed to be generated randomly in a sequence independent fashion, and have been termed 'Ignorant' sequences (Dover, 1980). Fincham (1983) refers to Ignorant DNA as ... "DNA which can proliferate differentially merely because of its repetitiveness rather than because of anything special about its sequence ..." and that ... "once a sequence acquires a certain abundance, in the form of tandem repeats, it may occasionally spread in almost epidemic fashion ...".

By applying the term 'selfish' or 'ignorant' to a sequence, it is possible to explain the presence and distribution of almost any component of a eukaryotic genome. Clearly then it is a term to be used with caution if the concept is not to be devalued.

### The maintenance of patterns and sequences

Although selection for specific sequences could be invoked to explain the maintenance of sequence homogeneity, the qualitative and quantitative variation between genomes of the majority of repetitive sequences might be interpreted as indicating that other mechanisms are operative.

Fixation by unequal crossing-over, a model discussed earlier, can account for the maintenance of relatively homogeneous tandem arrays of amplified sequences within a single linkage group and does not require that such DNA is subject to selective pressure (Smith, 1976).

This phenomenon might explain the organisation of the 400 bp satellite sequence in several Allium species. From data presented in Chapter 7 it appears as though the 400 bp monomer isolated from A. vineale hybridises to DNA from a number of different species; the pattern and degree of hybridisation however vary. To Bam HI digested DNA from A. sativum, A. babingtonii, A. porrum and A. schoenoprasum, hybridisation is mostly to a wide range of fragments, mostly of high molecular weight with no clearly discernable satellite 'ladder'. Conversely, to DNA from A. cepa, A. cepa var 'spring onion' and A. triquetrum, much of the hybridisation is to a ladder of fragments suggesting a degree of sequence maintenance.

The distribution and organisation of the 400 bp

family among the Allium species is difficult to explain in terms of differential amplification of sequences. Why related families should be present in A. cepa and A. vineale in an apparently well conserved organisation while being apparently rather heterogeneous within species closely related to A. vineale (e.g. A. sativum) is difficult to explain. However if homogenising mechanisms operate in some species and not others, the data can be explained.

According to Ved Brat (1965b) A. sativum and A. babingtonii are asexual, while A. cepa, A. triquetrum and A. schoenoprasum have sexual breeding systems. A. porrum has both systems. There appears to be some relationship between breeding system and the degree of maintenance of the 400 bp array.

If a tandem array was present in a sexually reproducing ancestral species it may have been maintained by unequal crossing-over and inherited by subsequently diverging species. If such species continued to maintain sexual breeding systems, homogeneous arrays might have been maintained by meiotic unequal exchange and still be visible as ladders to the present day. Fixation of variant sequences of the same repeat length might be responsible for the reduction in homology between species, as is seen between A. Cepa and A. vineale

for example. However in species which have acquired an asexual breeding system, unequal crossing over would perhaps not maintain homogeneity in the satellite arrays if it is a process operating only in meiosis as envisaged in the model of Smith (1976). Although sequences sharing some homology with the A. vineale 400 bp monomer probe would still be detectable in the present day genome they may have undergone much mutation and possibly reorganisation (especially if free of any selective constraint) - a situation perhaps evidenced by a smear of hybridisation in digests of A. sativum and A. babingtonii in figure 29.

The same distribution and organisation might of course arise if the satellite component, in particular perhaps its repeat length, was performing some function at meiosis. In sexually reproducing species, maintenance of the tandem arrays by selection for any meiotic function can be invoked.

Until hybridisation of pAv 1101 to a series of different digests of genomic DNA from a number of species has been performed, and until the breeding systems of the Allium species have been more clearly established, the above interpretation must be regarded as highly speculative. The correlation between breeding system and hybridisation pattern

is by no means perfect and the report that A. vineale is asexually reproducing (Ved Brat, 1965b) does nothing to strengthen the argument outlined above...

Nevertheless species of the Allium genus would appear to be suitable material for studying the requirement for meiosis in the maintenance of satellite DNA sequence homology.

Unequal crossing-over could also be responsible for the maintenance of sequence homology in tandem-arrays of genes. Selection alone would not be able to achieve this because the large numbers of genic sequences involved would dilute the effect of mutant copies. Selection would however play a role in preventing the fixation of deleterious mutations in genic sequences.

Repetitive sequences not subject to selection or to stochastic phenomena such as cross-over fixation might be expected to accumulate mutations, independently, from the moment of their generation. This ought to be particularly true of multiple copy dispersed sequences which cannot, due to their physical arrangement, be subject to unequal crossing over.

Reassociation work however seemed to suggest that dispersed repetitive elements of common origin shared more homology with one another than would be expected considering estimates of the rate of

mutation in the absence of selection. In cereal genomes for example it appeared as though older sequences were no more diverged than more recently evolved sequences, where age was estimated from evolutionary distribution (Flavell et al., 1977). Furthermore, the repeated sequences of a family appeared to be more closely related to each other within a species than to the corresponding family in other species. Such an observation was originally interpreted as being evidence for multiple amplifications of related but slightly diverged sequences in different genomes, the reamplification events belying the actual age of the sequences involved; in the light of more recent data however, other mechanisms have also been invoked.

Greater within-species than between-species homogeneity has been observed in a number of repetitive sequence families; in multigene families such as the rRNA genes (Coen et al., 1982; Strachan et al., 1982), and in dispersed non-coding families such as MIF-1 (Brown and Dover, 1981). Variation is accumulating in these families but within a species it appears to be accumulating in a 'concerted' fashion; individual members are not diverging in sequence independently. Although the mechanism is not clear, the process would seem to involve the replacement of a sequence in the genome



by a variant, a phenomenon which has been termed 'homogenisation'. The model of unequal exchange implies that such a process can only act within a tandem cluster of sequences; although it can maintain sequence homology within an array it cannot act inter-chromosomally. The homogenisation of dispersed sequences would seem to require a process which can operate between sequences on non-homologous chromosomes. Such a process might be 'gene-conversion', a non-reciprocal transfer of information. Evidence of gene conversion between tandemly arranged repeated genes in Saccharomyces has been demonstrated by Jackson and Fink (1981), and between dispersed repetitive sequences in yeast by Scherer and Davis (1980). The process in the latter case appeared to involve the formation of hybrid DNA and a correction of base pair mismatches. In the same organism, the replacement of a sequence by another as a result of a specific DNA transposition-substitution event appears to operate at the mating type locus, (Strathern et al., 1982).

Non-reciprocal gene conversion events have also been invoked to explain the sub-populations of MIF-1 sequences in the mouse genome (Brown and Dover, 1981). Any bias in the direction of conversion may result in the spread of variant

sequences throughout a genome, a hypothetical process termed molecular drive. Dover (1982) has proposed that molecular drive can be considered capable of distributing novel types of mutation by a process not dependent upon the external factors of natural selection or genetic drift. Whatever the mechanism behind this phenomenon, be it directionally or randomly by unequal exchange, gene conversion or direct transposition, it is proposed that it would be capable of spreading a variant sequence throughout a chromosome, between chromosomes and, at a higher level, throughout a sexual population (Dover, 1982; Dover et al., 1982). Theoretical considerations suggest that any directionality in a process such as gene conversion would indeed accelerate the rate of fixation of a sequence (Nagylaki and Petes, 1982) and that gene conversion would provide an efficient mechanism for maintaining the homogeneity of repeated sequences in eukaryotic chromosomes (Birky et al., 1976).

The evolutionary implications of such a model are considered later.

It should however be noted that much of the evidence for within-species homogeneity of members of repetitive sequence families has been provided by reassociation studies and, more recently, by restriction-mapping. DNA sequencing studies have

more recently established however that there is, in many instances, far more extensive variation within a species than previously suggested. Sequencing of cloned members of some satellite DNA families has revealed variation of the order of 12% in bovine 1.706 satellite for example (Miklos, 1982). Considerable heterogeneity has also been detected in the MIF-1 family (Brown and Piechaczyk, 1983, Dubnick et al., 1983) although certain regions of each element do appear to be more conserved than others.

The data presented here suggests that members of the B1000 family in A. sativum are somewhat heterogeneous in sequence as judged by restriction patterns and thermal elution profiles. The sample number is however too low to establish whether this is true of the majority of members of the family.

The sequencing of a few members of any large family will always be subject to sampling error. It might however be worth considering that the apparent homogeneity of some sequence families is misleading, and that further investigation will reveal more sequence heterogeneity than was previously suspected.

## CHAPTER 9

### **FUNCTIONS AND EFFECTS ATTRIBUTABLE TO REPETITIVE SEQUENCE DNA**

FUNCTIONS AND EFFECTS ATTRIBUTABLE TO REPETITIVE-  
SEQUENCE DNA

The question to which the majority of investigations into genome organisation (summarised in Chapter 1) were addressed was whether a study of the evolutionary history and present-day organisation of repetitive sequences would give any indication as to their function. How successfully has this question been answered?

As described earlier, patterns of genome organisation appear to be conserved in a wide range of eukaryotic species as do at least portions of the sequences of constituent elements in many cases (see Chapter 8 ). If selection is acting so as to preserve patterns and sequences it might be assumed that both are performing some function of advantage to the individual. If however the concepts of Selfish and Ignorant DNA, and stochastic models such as unequal crossing over and sequence homogenisation are accepted, there is no real need to invoke selection as a direct causative agent of all that is seen in the genome, much of the DNA therein may have no function in the true sense of the word. Nevertheless this does not mean to say that such DNA need have no effect upon the genome and/or individual; even if a family of sequences is generated by entirely random, stochastic processes

utterly unconnected with the functioning of the organism, it may still have consequential effects upon the genome.

What functions have been ascribed to repetitive sequences and what effects of their presence and turnover have been observed?

#### Attributable functions

##### (i) tandem arrays

The function of tandemly arranged multigene families provides few conceptual problems. With a few exceptions such as orphans and pseudogenes, most family members are functional genic sequences which are transcribed. Their reiteration may be a response to the requirement for large amounts of gene product.

The function (if any) of satellite DNA is by no means as clear. The fact that the vast majority of satellite families show no detectable levels of transcription appears to rule out any role as a transcriptional template. A role in the regulation of gene expression also seems unlikely considering the evolutionary variability in sequence and amount, and its heterochromatic distribution. It is in fact this association with heterochromatin - itself not a very clearly understood component of the genome - that has led to most speculation regarding any functions.

Satellite DNA has been implicated, with little evidence, as being responsible for recognition and pairing between homologous chromosomes and for the folding of chromatin within the chromosomes, so-called 'housekeeping functions' (Walker, 1968). The evidence for such functions is however not strong. The presence of satellite DNA does not seem to be essential for the condensation of facultative heterochromatin, (Brown, 1966) and although nucleosome phasing could be interpreted as being evidence of a structural role for satellite DNA in chromatin material (Musich et al., 1977) it could just as easily reflect one of the constraints upon the mechanism of its generation. Drosophila chromosomes deficient in most of their satellite DNA are still able to pair normally (Yamamoto, 1979) and yeast centromeres function at mitosis and meiosis without the massive amounts of satellite DNA normally associated with eukaryotic centromeric regions (Clarke & Carbon, 1980). Such evidence has led John and Miklos (1979) to conclude ...."the bulk of simple-sequence DNA can have very little to do with general cellular processes or with such concepts as centromere strengths..."

#### ii) Interspersed sequences

Specific functions have been proposed for dispersed sequences, for SINES in particular, although the construction and testing of models is complicated

by the fact that only a small subset of those sequences considered to be repetitive may be functional.

Repetitive sequences have been implicated in the control of gene expression, one of the fundamental problems in molecular biology, either as signal sequences in the DNA or as RNA processing signals. The early models of Britten and Davidson (1969) and Georgiev (1969) suggested for example that SINES acted as sites in the DNA for the specific binding of regulatory macromolecules, analagous to the promoters and operators of prokaryotic operons. Such models have been modified in the light of evidence relating to the organisation of SINES in the genome (Davidson et al., 1977) in an attempt to explain the co-ordinated control of unlinked genes. The vast majority of SINES do not however, it is fair to say, have the properties, such as conservation of sequence, expected of regulatory elements.

The extensive representation of interspersed repetitive sequences in cellular RNA has been interpreted as suggesting that such sequences are involved in the regulation of gene expression, perhaps as a result of the control of mRNA maturation (Davidson and Britten, 1979; Murray and Holliday, 1979; Lerner et al., 1980; Roberts, 1980).



Indeed some SINE families appear to be functional genes; cytoplasmic 4.5S RNA of rodent cells and 7 S RNA of human cells have both been implicated in the regulation of gene expression (Zieve & Penman, 1981; Lerner and Steitz, 1981). Both are transcripts of DNA sequences which can be regarded as SINES, being only a few hundred nucleotides in length.

It is however the presence of repetitive transcripts in heterogeneous nuclear RNA (HnRNA) and cytoplasmic poly(A) RNA that is of particular interest, especially as their expression is often developmentally modulated (Zuker and Lodish, 1981; Ryffel et al., 1981; Davidson and Posakony, 1982; Jelinek & Schmid, 1982 ; Posakony et al., 1983; Chung et al., 1983). A correlation between repetitive sequence expression and developmental change does not necessarily imply a functional role; such sequences may simply have transposed into transcription units, as seems to be the case for some repetitive elements in the sea urchin genome (Posakony et al., 1983). Unfortunately investigation is made difficult by the fact that only a small subset of the total number of elements in a repetitive sequence family is transcribed, the rest being transcriptionally inactive. With this in mind Davidson and Posakony (1982) conclude that ... "despite their ubiquity, their quantitative prominence, their apparent developmental regulation

and the amount of interest they have aroused, the repetitive sequence transcripts of animal cells remain a phenomenon in search of a physiological meaning ...".

The transcription of Bl000 sequences in A. sativum has not been conclusively detected (see page 69 ). It is however possible that under the conditions employed a small subset of transcriptionally active elements would not be detected.

A number of functions unconnected with gene expression have been proposed. The fact that 14 bp of the consensus Alu sequence shows homology with a number of viral origins of replication (Jelinek et al., 1980) has been interpreted as suggesting that Alu-like sequences act as origins of DNA replication in those genomes possessing them. Mammalian DNA is replicated bidirectionally simultaneously from many dispersed sites, approximately as many as there are Alu elements (Jelinek et al., 1980, Georgiev et al., 1981). There is currently insufficient evidence to conclude whether or not this is merely fortuitous.

Moderately repetitive DNA sequences synthesized during pachytene in Lilium (Bouchard and Stern, 1980) have been hypothesised as having a specific function in the process of chiasma formation and crossing-over. Such sequences (P-DNA) make up

approximately 1% of the genome, are of modal length 1500-2000 bp, are interspersed and are members of several distinct families. Members of each family appear to be non-divergent in sequence as judged by thermal stability analysis.

It has been suggested that the nicking and repair activities during pachytene occur at these specific P-DNA sites in the genome. Homologous P-DNA is also detectable by reassociation in other plant species across a broad phylogenetic range (Friedman et al., 1982). Although copy number varies considerably, the sequence of such DNA is strongly conserved; such conservation of sequence among moderately repetitive DNA elements of different species is very unusual, suggesting that there are strong selective forces ensuring the homogeneity of P-DNA.

Specific sequence recognition of a moderately repetitive component of a genome has also been proposed by Liu & Lark (1982). Deletions in  $\lambda$  clones containing a 3.25 Kb interspersed repeated DNA sequence from Dipodomys ordii appear to be the result of the recognition, by components of the  $\lambda$  Red (E. coli recE) function, of a specific sequence, and even of its polarity.

It is perhaps worth pointing out at this stage that care has to be taken when assigning functions to sequences which are considered repetitive. The

majority of the DNA hybridisation experiments relating to the organisation of repetitive DNA sequences in eukaryotic genomes have been performed using 'standard conditions', usually at an arbitrary  $T_m - 25^{\circ}\text{C}$ . in 0.12M PB which is consequently that stringency at which those sequences which have reassociated by a certain Cot value are termed repetitive and the remaining ones non-repetitive. As Bendich and Anderson (1977) point out, although this is the level at which investigators distinguish repetitive from non-repetitive sequences, there is no reason why it should be the same level of discrimination as applied by the organism.

The effect of temperature upon the re-association of repetitive sequences is clearly demonstrated for example by the data presented in figure 24. At lower stringency, genomic sequences in A. cepa showing some homology to B1000 are detectable; such sequences are not detectable on the autoradiograph at 'standard' and 'high' stringency. Similarly at higher temperatures there is reduced hybridisation of B1000 to genomic DNA of A. sativum and A. porrum and the pattern of hybridisation is changed; bands of presumably relatively well matched sequences become more prominent over the background smear. Thus there appear to be many

sequences which although considered homologous to B1000 at 65°C. are not detectable as being so at 75°C., suggesting a degree of divergence of sequence among B1000 related sequences.

By classifying many rather heterogeneous repeats as repetitious, it is quite possible that a few very homogeneous repeats are being obscured. If it is only the latter that is of functional significance to the cell then different experimental criteria will have to be adopted in order to distinguish such sequences from the 'noise' of the heterogeneous background sequences. Few studies have however been performed at other levels of stringency.

Another danger associated with molecular studies is that of a parochial approach to DNA sequence alone. Manuelidis (1982) stresses the importance of determining the configuration and organisation of repetitive elements at a higher level than nucleotide sequence. Suggesting that such sequences may have an influence upon higher order three dimensional architecture of chromosomes, Manuelidis (op.cit.) has shown that members of the c-Hind R family of interspersed repeated sequences are non-randomly distributed on human chromosome arms and may participate in the alignment and/or association of chromosomes.

The association in interphase nuclei of specific regions of non-homologous chromosomes that are rich in particular types of DNA sequences might suggest a function for such sequences in chromosome recognition. In this context Bassi et al. (1982) have demonstrated that highly repetitive inverted-repeat DNA appears to aggregate in interphase nuclei of Vicia faba and that metaphase chromosomes are preferentially labelled in central regions after in situ hybridisation with labelled inverted-repeat DNA. In A. cepa however such sequences appear to be scattered throughout all chromosomes (Stack and Comings, 1979).

Chromosomes of A. cepa do appear to adopt a 'telophase configuration' throughout interphase with centromeres lying at one side of the nucleus and telomeres at the opposite side (Fussell, 1975). These regions are identifiable in interphase nuclei because they contain late replicating DNA. The nature of such DNA has however yet to be established.

As stated earlier, the presence of a particular sequence component by no means requires it to have a function in the genome; selfish and ignorant DNA require no function to be invoked in order to explain their presence in the genome. The danger of avoiding the question of function by simply characterising a sequence as selfish or ignorant was alluded to earlier however.

Whether serving a function or not, the physical presence of repetitive DNA sequences may have consequences; a number of effects (both specific and general) have been characterised.

#### Attributable effects

It would be surprising if a DNA sequence of length several hundred to several thousand base pairs, present in as many as several thousand copies and possessing the ability to transpose was observed to have no effect upon the genome - even if such a sequence was regarded as 'functionless'.

Indeed many phenomena have been described that appear to be the direct result of repetitive sequences. Some are very specific effects, others the more general effects of DNA accumulation.

Insertion-mutations caused by transposable elements have been characterised in yeast (Williamson et al., 1983), in Drosophila (Modolell et al., 1983) and in Soybean (Goldberg et al., 1983). Such elements are assumed to be responsible for a number of genetic instabilities, such as ones in Maize (McClintok, 1956) and mouse (Whitney and Lamoreux, 1982).

In addition to being the cause of some insertional mutations, SINES and LINES appear to be responsible for other types of major rearrangements of DNA in the genome. A large deletion responsible

for a disorder of haemoglobin synthesis in humans has one end point half-way into an Alu sequence (Jagadeeswaran et al., 1982). This of course may be coincidental, although Calabretta et al. (1982) report the presence of genetic instabilities in areas of the human genome containing Alu sequences. Furthermore the association of a long Drosophila transposable element with chromosomal deletion breakpoints has been reported by McGinnis et al. (1983) and Perlman (1983).

It may be that the facilitation of sequence rearrangements by repetitive sequences is a common phenomenon in eukaryotic genomes.

Both tandemly arranged and interspersed repetitive sequences also appear to have an effect at a level higher than that of nucleotide sequence rearrangement.

A well documented effect of satellite DNA which may be of great evolutionary significance, is the reduction in the frequency of recombination events occurring within adjacent sequences during meiosis in Drosophila; it appears to be the amount of satellite DNA that determines this germ-line effect, the sequence itself seems relatively unimportant (John and Miklos, 1979; Miklos et al., 1980; Miklos and Gill, 1981; Miklos, 1982).

Of particular interest therefore is the



relationship between chiasma frequency and variability in some plant species. Variability is lower in populations of Lolium and Festuca with high chiasma frequencies and vice versa, while selection appears to be more effective among populations with low chiasma frequencies (Rees and Jones, 1977). Rees and Jones (op.cit.) conclude that ... "the information .... provides strong evidence of the adaptive consequences of adjustments in chiasmata upon the variability of populations and their capacity to respond to selection."

Any effect upon recombination frequency by satellite DNA might therefore be of great significance and might imply that the presence and quantity of a satellite component of a genome is subject to selection to a greater degree than previously imagined. There are however no data yet relating to whether any effects of satellite DNA are of adaptive significance; whether it can be regarded as having a "germ line function" (John and Miklos, 1969) is open to interpretation.

As Maynard-Smith (1982) points out, "Usually the distinction between an effect and a function is fairly clear, but difficulties arise when the effect is not on the individual, but on the species..."

The majority of other 'general effects' of genomic repetitive sequence components are ones due

to quantity rather than nucleotide sequence.

Although in the original models selfish and ignorant DNA is assumed not to be subject to phenotypic selection, the amplification processes might result in changes in quantity of DNA. Such changes, many authors argue, could be subject to selection, although the point at which selection begins to act, and its subsequent intensity, is a matter of some debate.

Cavalier-Smith (1978) points out a strong positive correlation between C-value and cell-size and also length of mitotic and meiotic cycles. Consequently there appears to be an inverse correlation between C-value and growth rate, and hence between C-value and development time. He concludes that C-values are highly adaptive and controlled by a balance of selective forces. If this is true then presumably any changes in DNA content - by amplification or deletion of sequences - will not be selectively neutral. That C-value determines the range of life-cycle types that plants can display in particular environments has also been stated by Bennett (1977, 1982), while Grime & Mowforth (1982) present evidence that genome size is related to life-cycle type in British flora.

The effect that nuclear DNA has upon the phenotype of an individual as a result of its quantity

has been termed "nucleotypic" (Bennett 1977), a term that describes non-genic characters that affect phenotype by bulk rather than by informational content. Nucleotypic effects are apparently not confined to the effects upon cell size or development rate. Bennett (1982) suggests that nucleotypic effects can determine chromosome architecture and spatial ordering of haploid chromosome sets.

The disposition of a chromosome in the haploid genome appears to be determined by the size of its arms (which in turn is determined by the arms DNA content), arms of similar length associating in a predictable fashion in interphase nuclei (Bennett, *op.cit.*, Heslop-Harrison and Bennett, 1983 a & b). This differs from previous chromosome recognition models in that it predicts that associations are due to the quantity of DNA on each arm and not to the nature of the sequences therein.

Differences in C-values between Allium species may be due to selection acting upon the nucleotype. If chromosome association is required for the correct behaviour of chromosomes during mitosis for example it might be predicted that an optimal configuration will be maintained by selection. If the configuration is determined by relative arm lengths then, if the chromosomes of a haploid set were to gain DNA by whatever mechanism, they might be expected to gain it in a concerted fashion so as to maintain the optimal configuration.

An amplification event in one chromosome might result in selection favouring a similar event in an associating chromosome of the haploid set. If the sequence subject to amplification is of little consequence then different segments might be amplified at each amplification event, a situation which could explain the large differences, between species, in the nature of moderately repetitive families.

The problems associated with work on Allium species makes this a difficult hypothesis to test; the taxonomy of the genus is somewhat ambiguous and phylogenetic relationships between species are not at all clear. Furthermore the difficulty in distinguishing between many of the chromosomes of

the complements of many Allium species (see figures 4 & 5 ) renders association studies somewhat difficult. There is some evidence that there are associations between the chromosomes of a haploid set in the genome of A. sativum, associations being between chromosome arms of similar size (G. Butler pers.comm.) but analysis of the data is as yet incomplete.

In the context of the nucleotypic effects of DNA mass, it is interesting to speculate on the constraints (if any) to DNA accumulation within a genome. The rather strong selective constraints envisaged by Cavalier-Smith (1978) were described earlier, as were the constraints imposed by the contribution of DNA quantity to chromosome disposition (Bennett, 1982).

In some situations where DNA accumulation during evolution appears to have occurred, DNA amounts within related species appear to be discontinuously distributed.

In Lathyrus species for example, DNA-content values are clustered into groups separated by intervals of about 4 pg, a situation also found to exist in the Allium genera (Narayan, 1982). This might represent a series of steady states into which species must fall, quantum changes being responsible for species evolving into higher states. The addition of DNA during evolution of the

Lathyrus species appears to be approximately equal on all chromosomes of each complement (Narayan and Durrant, 1983) with the result that there is constancy in chromosome number, chromosome shape, karyotype and in the ratio of repetitive to non-repetitive DNA during evolution. DNA changes in Allium species also appear to be widely distributed between chromosomes (Jones and Rees, 1969). Evidence of a similar situation in the animal kingdom comes from some amphibian species. The karyotypes of 26 Plethodon species are identical except for the amount of DNA per haploid set (Macgregor, 1982), suggesting that chromosomes have grown in size due to the accumulation of DNA, in a balanced, proportional manner. 'Karyotype selection' may, in such a way, be responsible for the present day physical organisation of DNA sequences in the genomes of different species. There may well be constraints upon changes in genome organisation during evolution, but which factors determine such constraints are not clear.

The variation in the repetitive sequence component of genomes, in the nature of the repeat families and their number, might be determined by the capacity of a particular genome to generate repeated sequences (or to be 'infected' by them)

and the capacity of a genome to tolerate such change. Bouchard (1982) argues that a combination of such factors will determine the physical organisation of a particular genome.

In this model plant genomes would be examples of genomes possessing a high level of predisposition to the generation of repeated sequences, in addition to possessing a high tolerance to accumulation of DNA sequences. Such a situation would be expected to lead to a large genome dominated by a rather heterogeneous moderately repetitive sequence component. Such an organisation is typical of a large number of plant genomes, including those of Allium species.

If, in a number of genera, karyotypes are maintained by some force so that the only gross difference between species is that of DNA amount, does this imply that DNA accumulation can contribute to speciation events, either as a consequence of the increase in quantity or of the change in sequences?

## CHAPTER 10

### REPETITIVE DNA AND THE EVOLUTION OF SPECIES



### REPETITIVE DNA & THE EVOLUTION OF SPECIES

Progress in studies relating to the molecular organisation of eukaryotic genomes has raised the question of whether changes in the repetitive sequence DNA components of chromosomes can be causative agents in the phenomenon of speciation, described by Fincham (1983) as "the most interesting but also the most inaccessible area of evolutionary study ...".

While it is widely accepted that the key feature in speciation is the attainment of reproductive isolation, such an attainment may be due to any one of many initial factors, or indeed probably to a combination of several. Such factors are classically considered to be for example prezygotic mechanisms such as isolation (e.g. ecological, temporal, ethological, mechanical) or post-zygotic mechanisms such as hybrid inviability or sterility.

Speciation is often accompanied by a number of genetic and cytogenetic changes. Distinct but related species often show differences in alleles as detected by gel electrophoresis. At a higher level, they very commonly show differences in karyotype; this may be in the form of a change in chromosome number, major structural changes such as translocations, fusions and inversions within a chromosome complement, a change in DNA amount or, in

the case of many plant species, polyploidisation. Changes in karyotype have clearly occurred during the evolution of Allium species as described earlier.

All the aforementioned have been implicated as causative agents in speciation but whether this is indeed the case or whether such changes have merely accumulated following the speciation event is not clear. Nor is it easily testable despite the fact that different components of the eukaryotic genome can now be isolated and examined in some detail. Separating the causative and associative events of speciation will necessarily remain frustratingly difficult as all studies must be retrospective.

The discovery and characterisation of repetitive sequence components from a wide range of organisms has permitted the evolutionary history of many such sequences to be proposed, some of which were discussed earlier. Both the rate of sequence turnover and the magnitude of its effects upon the physical structure of a genome were unsuspected when the majority of the classical models of speciation were proposed. Consequently with its discovery came many attempts to incorporate the behaviour of repetitive sequences into evolutionary theory.

On the premise that the component of the genome which is most different between two species is a probable source of the initial speciation event, much

early work focussed upon satellite DNA. The quantity, relatively homogeneous nature and very localised distribution of satellite DNA sequences within a genome suggested a role in meiotic pairing. If so, the species specific organisational patterns and the apparent quantitative and qualitative changes over short evolutionary periods would seem to imply an evolutionary role for satellite DNA sequences in the creation of mating-barriers (Fry and Salser, 1977). A failure to demonstrate a role in meiotic pairing for these sequences has however rather invalidated this argument.

Hatch et al. (1976) claim that a satellite component of Dipodomys species is responsible for major chromosomal rearrangements and can thus be involved as a speciation-inducing agent. Again the case remains unproven.

A review of the available literature at the time has been presented by John and Miklos (1979) in which they conclude that there is no evidence implicating satellite DNA in speciation events. Data published since that time lend support to this (Miklos and Gill, 1981 ), White (1982) concluding "...on present evidence we should exclude changes in satellite DNA as primary causes of speciation ...".

Attention has more recently been turned towards the moderately repetitive DNA sequence component of

the eukaryotic genome, the characteristics and evolutionary behaviour of which were outlined earlier (see Chapters 1, 8 and 9).

The effects of moderately repetitive sequences, or perhaps more particularly the effects of their 'turnover', upon a genome and their role in the establishment of barriers to gene exchange between populations remain somewhat speculative.

There is now however a well documented example in which repeated DNA sequences have been directly implicated as the cause of reproductive isolation between certain Drosophila strains, the phenomenon being 'hybrid dysgenesis', the mechanism the so-called P-M system. Certain strains of Drosophila when interbred produce progeny that are dysgenic, displaying traits such as sterility and high rates of recombination, mutation and chromosome aberrations. Many of the mutations are unstable, there being a high rate of reversion to wild-type. It has been demonstrated (Rubin et al., 1982) that such dysgenic traits are due to the insertion and precise excision of transposable 'P-elements', the activation of such elements being dependent upon the cytoplasmic background of the chromosomes harbouring them. In a cytoplasmic P background such elements appear to be stable, transposition being repressed. In an M cytotype however P-elements, of which there are about

30 dispersed throughout the D.melanogaster genome for example, become de-repressed and highly unstable. 'Contamination' of chromosomes by P-elements appears to be rather rapid suggesting that the hybrid sterility caused by the dysgenic effects of the P-M system might result in the reproductive isolation of populations geographically isolated for as little as 10 years (Rubin et al., op cit.). Rose & Doolittle (1983) have named this transposable element dependent model of speciation the "genomic disease" model. Although the presence of such dysgenic systems with associated pleiotropic genetic effects in organisms other than Drosophila has yet to be demonstrated, the existence of transposable-element like sequences in a wide range of other species (see Introduction) might be taken to imply that the 'genomic disease' model of speciation could apply to a number of natural situations.

In addition to 'genomic disease', Rose & Doolittle (1983) identify two further models of speciation involving repetitive sequences.

The first, 'mechanical genome incompatibility' does not require the activation of transposable elements, merely requiring that alterations in the organisation of repetitive sequences, perhaps by amplification or homogenisation, affect chromosome interactions. A failure of chromosomes to pair during meiosis could lead to meiotic non-disjunction during the formation of gametes, resulting in post-zygotic

reproductive isolation. It might be hypothesised that such a situation could occur as a result of gross structural differences between chromosomes, which in turn might result from massive amplification of sequences in one of the two interacting genomes only. The situation is however not as simple as this as pairing has been observed between homeologous chromosomes which differ considerably in length and DNA content. An example from the Allium species is in the hybrids produced by an A. cepa x A. fistulosum cross (Jones & Rees, 1969). Loops and overlaps occur in bivalents as a result of the chromosome volume (and DNA content) of the A. cepa complement being some 30% larger than that of the A. fistulosum complement. A similar situation has been described in some Lolium species hybrids in which homeologous chromosomes are capable of complete and effective pairing despite large differences in the DNA content of each complement (Seal and Rees, 1982).

On the other hand some good species have been observed to possess what appear to be visibly identical chromosomes to those of a related species, for example the homosequential species of Hawaiian Drosophila (Carlson et al., 1969).

Such examples suggest that meiotic chromosome pairing is determined by a small number of sites within the chromosome, a subset of the sequences

resident within. Consequently, an alteration in copy number or in the sequence of DNA at such sites as a result of repetitive sequence turnover would also lead to disruption of pairing. Stebbings (1982) has referred to such changes as 'submicroscopic structural differences' and states that "the origin of the majority of plant species is associated with chromosomal repatterning".

Nor need only meiotic mechanisms be affected by alterations to chromosome interactions. Interactions between non homologous chromosomes of a haploid set in somatic nuclei have been reported in plants and animals (Bennett, 1982; Manuelidis, 1982, Heslop-Harrison and Bennett, 1983 a,b ), associations apparently occurring between similarly sized arms of chromosomes.

If one chromosome complement were to experience amplification and/or redistribution on a scale large enough to disrupt such associations in a hybrid, the fitness of a hybrid individual might be affected. In this way a potential speciation-promoting mechanism can be invoked. Again, the problem of separating cause and effect prevents the conclusions that repetitive sequences are alone responsible for chromosome incompatibility.

The third model of molecular speciation is different from the previous two in that it requires

that at least some of the repetitive DNA sequences involved are functional. The 'genome resetting' model (Rose and Doolittle, 1983) suggests that changes in the position and sequence of repetitive elements, which are themselves regulatory elements for genes for example, could grossly affect gene regulation and development. If some repetitive sequences do function in a way suggested by Britten and Davidson (1969), reprogramming of development and differentiation leading to speciation might be the result of changes in the physical organisation of such sequences.

The evidence for the involvement of specific repetitive DNA sequences in the regulation of gene expression is, however, not strong. This merely enforces the view that before any conclusions regarding the role of repetitive DNA sequences in speciation can be drawn, much more needs to be known regarding the function of such sequences and their effect upon chromosome architecture and interactions.

There is clearly a large proportion of many eukaryotic genomes consisting of sequences whose rate of change of frequency and/or organisation is dramatically different from the rate of change determined for nucleotide substitution in protein coding sequences.



There is possibly therefore great potential within non-genic sequences for the relatively rapid generation of genetic diversity, particularly if, as seems likely, the reorganisation of repetitive elements can have pleiotropic effects, unlike a nucleotide substitution which may only affect a single gene.

Rapid, dramatic, temporally irregular events being responsible for macro-mutations which may result in speciation events is by no means a new concept. Alluded to by Goldschmidt (1940) it is an idea expanded more recently by Gould and Eldridge (1977) in their 'punctuated equilibria' model and by Stanley (1979) in his 'rectangular evolution' model. The problem with such evolutionary models is that very special circumstances have to be invoked in order to explain the initial survival of a grossly novel phenotype in the population of origin.

Data relating to the evolutionary history and present day organisation of repetitive DNA families in eukaryotes may be of relevance at this point.

Dover et al. (1982) have proposed that the

gradual homogenisation of a family of sequences could result in a synchronous change in the genomes of individuals in a sexually reproducing population; they would be subjected to a 'cohesive mode of evolution'. If the process were sufficiently slow, no significant increase in variance and therefore no significant change in relative fitness between individuals within a population would occur as a result of the homogenisation process.

If changes in the sequence of repetitive DNA families were to affect in some way the reproductive or general biology of the organism, then it is proposed that an effective reproductive barrier might be built up between those organisms carrying the fixed variant sequence and those not. The concerted spread of sequences through one population might in fact only be evidenced upon hybridisation with a second population from which the fixed sequences were absent. Hybrid dysgenesis might be the immediate result of such a cross, being replaced or complemented perhaps by further barriers to reproduction. Barriers such as adaptive morphological or allelic differences could subsequently result in an irrevocable speciation event.

The significance of molecular drive as a mode of speciation would presumably depend upon the nature of the sequences undergoing homogenisation, their effect upon the organism, and the rate at which the

homogenisation mechanism operated; herein lies sufficient as yet inaccessible data to maintain lively argument between proponents and opponents of the theory.

One mechanism of speciation is not in doubt. Polyploidisation, a macromutational event almost exclusive to plant species can result in the establishment of a separate species in a single generation and is clearly documented in a vast number of plant genera, including Allium (see Table 1 ). One of the best documented examples is in the phylogeny of wheat, an allohexaploid derived from three diploid genomes (summarised by Rees and Jones, 1977). Polyploidisation (either autopolyploidisation or allo polyploidisation) can be considered a special case in that there is no differential replication of a particular component of the genome but the effects of an increase in DNA content might be similar to those considered earlier for amplified repetitive sequence components.

In conclusion, despite the fact that repetitive sequence turnover in the eukaryotic genome is an apparently common phenomenon, it has yet to be adequately incorporated into evolutionary theory. Until causative and associative effects can be separated it is difficult to know where to look for evidence. With the discovery of repetitive sequence

turnover and the realisation of its implications ...

"it is now much more difficult ..." as White (1982)

concludes ... "to formulate universal laws or

principles with regard to any evolutionary process ...".

## **CHAPTER 11**

### **CONCLUSIONS AND SPECULATIONS**

### CONCLUSIONS AND SPECULATIONS

Species of the genus Allium would seem to be good material upon which to base investigations relating to the nature of differences between species in repetitive DNA sequence components. Clearly a large proportion of some Allium genomes comprises repetitive sequences and there is considerable variation in C-values between even rather closely related species (see Table 1).

The nature of the organisation of repetitive DNA sequences in plant genomes has been the subject of surprisingly few investigations; the majority of these have focussed upon tandemly organised heterochromatic DNA. The work presented here forms part of the initial investigations into the organisation of Allium genomes and into the nature of the differences between such genomes.

The satellite component of A. vineale discussed in Chapter 7 is, along with that reported by A. James and S. Barnes (pers.comm.), the first report of satellite DNA in any Allium species. In addition however data is presented which represents one of the first attempts to characterise an apparently non-tandemly organised DNA sequence family in a plant genome. Members of this B1000 family appear to be interspersed with other repetitive sequences in the genome of A. sativum, probably as part of larger repetitive elements which are somewhat heterogeneous in sequence, have a

reiteration frequency of at least  $10^4$  and do not appear to be abundantly transcribed. In these respects B1000 members appear to conform to the definition of LINE sequences as discussed earlier (page 26 ). Sequences sharing some homology with a representative cloned B1000 member are detectable in the genomes of many other Allium species; their reiteration frequency and/or sequence divergence is however very variable, as is in many cases their organisation.

The organisation of B1000 members at the nucleotide level is in all species clearly rather complex. Hybridisation of the cloned member to digested total DNA of several species results in a number of discrete bands of a variety of intensities, none of which seem to bear any simple numerical relationship in size with any others (Chapters 4, and 6). The majority of the explanations for the present day organisation of eukaryotic interspersed sequences, (reviewed in Chapter 8), could be applied, at this early stage in the story, to the B1000 sequence family. It is impossible to say yet whether such sequences, or more particularly the elements containing them, are capable of transposition; their distribution however could be interpreted as suggesting this. Amplification of unit sequences certainly seems to have been part of the evolutionary history of the B1000

family. Presumably the bands of hybridisation mentioned above are a result of relatively recent amplification events - a segment of DNA containing at least part of an ancestral B1000 sequence might be amplified and subsequently accumulate mutations with time. The loss of homology with increasing number of mutations would explain the absence of hybridisation to the smear of diverged sequences at higher stringencies in figure 24 for example.

Transposition and reamplification events would add to the complexity of the situation, as would any process of homogenisation. The data obtained so far is too complex to enable the contribution of such phenomena to be clearly established.

A study of the organisation of fragments other than of length 1050 bp to which B1000 hybridises might shed some light upon the evolutionary history of the family. Preliminary results using a cloned 2.9 kb A. sativum DNA fragment, one of the fragments in the 2.9kb band highlighted by B1000 in Bam HI digest of total DNA (see figure 6) confirm the complexity of the situation.

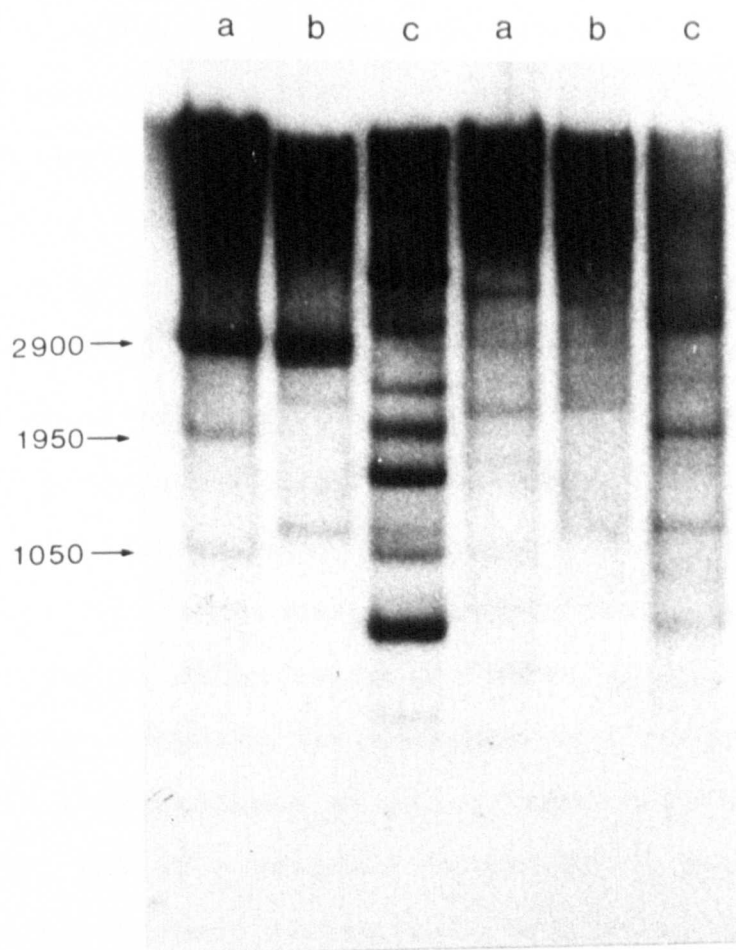
Figure 33 shows an autoradiograph resulting from the hybridisation of this 2.9 kb probe to various digests of A. sativum and A. vineale DNA. In the Bam HI digests, a band of 2.9 kb is highlighted, the length of the probe sequence. In addition there is, as with the B1000 probe, hybridisation to other discrete bands both larger and smaller than the probe length, and to a background



FIGURE 33

Total DNA from A. sativum (A.s.) and A. vineale (A.v.) digested with (a) Bam HI (b) Eco RI and (c) Hind III, fractionated, Southern blotted and hybridised with  $^{32}\text{P}$  nick-translated 2.9 kb DNA from an A. sativum Bam HI digest (see figure 6(a)).

The size of some of the major bands of hybridisation in the A. sativum Bam HI digest are marked in base pairs.



∪  
A.s.

∪  
A.v.

smear. Among these bands is one of length 1050 bp in the Bam HI digest, presumably fragments related to the original Bl000 probe.

As with the Bl000 probe, the hybridisation patterns are different between the two species. There are however bands highlighted by the 2.9 kb probe which are not highlighted by Bl000 (see figures 21, 22 and 23) - these being presumably due to repetitive sequences in the larger probe that are unrelated to Bl000. The situation is clearly rather complex. The elucidation of the nucleotide organisation in a number of larger Bl000-related DNA fragments might simplify it.

The fact that sequences related to the A. sativum Bl000 probe can be detected in all other Allium species so far examined suggests its presence in a progenitor species. Is this evidence of its maintenance throughout evolution and of a possible function? A degree of sequence maintenance throughout recent evolutionary history, as seems to be the case with for example P-DNA of Lilium (Friedman et al., 1982) would perhaps suggest a sequence specific function. There is however no such sequence maintenance within the bulk of Bl000 sequences, there apparently being a great deal of sequence divergence both between and within species (Chapters 5 and 6).

Does the genomic organisation of Bl000 in the different species provide any clues as to the

evolutionary history of the family? Clearly there are differences between species in those sequences in which B1000 family members are embedded; this is evidenced by the differences between species, albeit sometimes minor, in the patterns of hybridisation in for example figures 21, 22 and 23. It is interesting to speculate (and practically impossible to test) that such differences are the causative agents in the speciation events themselves (see Chapter 10). More likely though is that the maintenance of a sequence organisation like that of A. sativum in some species and the reorganisation in others probably reflects the stochastic mechanisms of repetitive sequence turnover to which Allium genomes have been and are presently being subjected.

Closely related species, only relatively recently diverged, might be expected to possess similar organisations of B1000. This seems to be true in the cases of, for example, A. babingtonii and A. porrum. Any differences are presumably due to reorganisation events (such as reamplification) subsequent to the species in question diverging from the A. sativum line of descent. Such events might be of relatively large magnitude, as seems to be the situation with A. vineale for example.

The further separated from A. sativum a species, the more might its B1000-related sequences be diverged from the cloned B1000 representative probe, as a

result of the accumulation of mutations along both lines of descent or as a result perhaps of the homogenisation of different variants in each species (if indeed such a process operates). Thus B1000-related sequences from the more distantly related species such as the cepa group not surprisingly display a large degree of sequence divergence from those of A. sativum.

Although more evidence is required, it seems that if the majority of B1000 family members do perform a function in the genome, it is presumably not sequence related. More likely however is that the B1000 family is merely part of the bulk of repetitive sequences which are subject to selection as a result of their nucleotypic effects (see Chapter 9).

Clearly, future work could proceed along many lines. Isolation of B1000-containing clones from a  $\lambda$  or cosmid vector library would be the first step in establishing whether a large conserved LINE element does exist. Sequencing of such elements might subsequently give clues as to their origin, although the apparently large degree of heterogeneity might render sequence relationships hard to establish. Detection of transcription of such elements might be indicative of the existence of transposable elements although if only a small subset are transcriptionally active this might not prove easy.

At a higher level it would be interesting to establish whether B1000 family members are clustered in the chromosome complements of Allium species; if so functions in chromosome recognition might be invoked. Such a study would be particularly interesting if metaphase chromosome association could be established; localisation of B1000 family members in particular segments might be indicative of a nucleotypic role such as a 'balancing' of the chromosome arms in a complement by amplification of DNA.

There has been a neglect of the plant kingdom in detailed genome organisation studies. It would be foolish to assume that plant genomes are any less complex than those of animals but perhaps not unreasonable to imagine that they are in many ways different in their organisation and functioning. The work presented here represents the early stages of investigations into the organisation of specific repetitive-DNA sequence families in species of the genus Allium.

Progress in the characterisation of repeated DNA sequences of eukaryotic genomes has in the last decade been considerable. The prospects for such investigations are perhaps best encapsulated by Bouchard (1982):

"It is quite conceivable that some new and unforeseen observations or the accumulation of enough detailed genome organisation and specific

sequence information will again reveal aspects of unity and simplicity underlying the apparent diversity which now seems to characterise moderately repetitive DNA in evolution.

"At present, however, it seems equally likely that further analyses will reveal more diversity .... it may well be that in considering moderately repetitive DNA, we confront a point at which the unity of biochemistry and the diversity of organic evolution meet ...".

# REFERENCES

- ANDERSON, D.M., SCHELLER, R.H., POSAKONY, J.W., MCALLISTER, L.B., TRABERT, S.G., BEALL, C., BRITTEN, R.J., DAVIDSON, E.H. (1981). Repetitive sequences of the Sea Urchin genome. Distribution of members of specific repetitive families. J.Mol.Biol. 145, 5-28.
- BARTELS, D., THOMPSON, R.D. (1983). The characterisation of cDNA clones coding for wheat storage proteins. Nucl.Acids Res. 11, 2961-2977.
- BASSI, P., CREMONINI, R. & CIONINI, P.G. (1982). Cytological localization of inverted repeated DNA sequences in Vicia faba. Chromosoma 85, 453-459.
- BEDBROOK, J.R., & GERLACH, W.L. (1980). Cloning of repeated sequence DNA from cereal plants. In: Genetic engineering Vol 2 (Setlow, J.K. & Hollaender, A., eds), pp. 1-19, Plenum Press, New York.
- BEDBROOK, J.R., JONES, J., O'DELL, M., THOMPSON, R.D., FLAVELL, R.B. (1980a). A molecular description of Telomeric heterochromatin in Secale species. Cell 19, 545-560.
- BEDBROOK, J.R., O'DELL, M. & FLAVELL, R.B. (1980b). Amplification of rearranged repeated DNA sequences in cereal plants. Nature, 288, 133-137.
- BELL, G.I., PICTET, R. & RUTTER, W.J. (1980). Analysis of the regions flanking the human insulin gene and sequence of an Alu family member. Nucl. Acids Res. 8, 4091-4109.
- BENDICH, A.J. & ANDERSON, R.S. (1977). Characterisation of families of repeated DNA sequences from four vascular plants. Biochemistry 16, 4655-4663.
- BENNETT, M.D. (1977). The time and duration of meiosis. Phil. Trans. Roy. Soc. ser. B 277, 201-226.
- BENNETT, M.D. (1982). Nucleotypic basis of the spatial ordering of chromosomes in eukaryotes and the implications of the order for genome evolution and phenotypic variation. In: Genome Evolution (Dover G.A. & Flavell, R.B., eds), pp. 239-261, Academic Press, London.



- BENNETT, M.D. & SMITH, J.D. (1976). Nuclear DNA amounts in angiosperms. Phil. Trans.Roy.Soc. ser.B. 274, 227-274.
- BIRKY, C.W., & SKAVARIL, R.V. (1976). Maintenance of genetic homogeneity in systems with multiple genomes. Genet. Res. (Camb.) 27, 249-265.
- BIRO, P.A., CARR-BROWN, A., SOUTHERN, E.M., & WALKER, P.M.B. (1975). Partial sequence analysis of mouse satellite DNA: evidence for short range periodicities. J.Mol.Biol. 94, 71-86.
- BLIN, N., WEBER, T. & ALONSON, A. (1983). Cross reaction of SnRNA and an Alu-I-like sequence from rat with DNA's from different eukaryotic species. Nucl. Acids Res. 11, 1375-1388.
- BONNER, T.I., BRENNER, D.J., NEUFELD, B.R. & BRITTEN, R.J. (1973). Reduction in the rate of DNA reassociation by sequence divergence. J.Mol.Biol. 81, 123-135.
- BOSTOCK, C.J. & TYLER-SMITH, C. (1982). Changes to genomic DNA in Methotrexate-resistant cells. In: Genome Evolution(Dover G.A. & Flavell, R.B., eds), pp. 69-93. Academic Press, London.
- BOUCHARD, R. A. (1982). Moderately repetitive DNA in Evolution. Int.Rev.Cytol. 76, 113-193.
- BOUCHARD, R. A. & STERN, H. (1980). DNA synthesized at Pachytene in Lilium: A non-divergent sub-class of moderately repetitive sequences. Chromosoma 81, 349-363.
- BOUGOURD, S.M. & PARKER, J.S. (1975). The B-chromosome system of Allium schoenoprasum L. B-distribution. Chromosoma 53, 273-282.
- BRITTEN, R.J. & DAVIDSON, E.H. (1969). Gene regulation for higher cells: A theory. Science 165, 349-357.
- BRITTEN, R.J., GRAHAM D. E. & NEUFELD, B.R. (1974). Analysis of repeating DNA sequences by Reassociation. In: Methods in enzymology Vol.XXIX(Grossman, L. & Moldave, K., eds), pp.363-418, Academic Press, New York.
- BRITTEN, R.J. & KOHNE, D.E. (1968). Repeated sequences in DNA. Science 161, 529-540.

- BROWN, S.D.M. (1983). A mouse dispersed repeat sequence showing remarkable similarities to the long terminal repeats of retroviruses. Gene 23, 95-97.
- BROWN, S. D. M. & DOVER, G. (1981). Organisation and evolutionary progress of a dispersed repetitive family of sequences in widely separated rodent genomes. J.Mol.Biol. 150, 441-466.
- BROWN, S.D.M. & PIECHACZYK, M. (1983). Insertion sequences and tandem repetitions as sources of variation in a dispersed repeat family. J.Mol.Biol. 165, 249-256.
- BROWN, S.W. (1966). Heterochromatin. Science 151, 417-425.
- CALABRETTA, B. ROBBERSON, D.L., BARRERA-SALDANA, H.A. LAMBROU, T.P. & SAUNDERS, G.F. (1982). Genome instability in a region of human DNA enriched in Alu repeat sequences. Nature 296, 219-225.
- CALLAN, H.G. (1967). The organisation of genetic units in chromosomes. J.Cell.Sci. 2, 1-8.
- CALOS, M.P. & MILLER, J.H. (1980). Transposable elements. Cell 20, 579-595.
- CAMERON, J.R., LOH, E.Y., & DAVIS, R.W. (1979). Evidence for transposition of dispersed repetitive DNA families in yeast. Cell 16, 739-751.
- CARSON, H.L., CLAYTON, F.E. & STALKER, H.D. (1967). Karyotypic stability and speciation in Hawaiian Drosophila. P.N.A.S. 57, 1280-1285.
- CAVALIER-SMITH, T. (1978). Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. J.Cell.Sci. 34, 247-278.
- CHEN, M. SHIMADA, T., MOULTON, A.D., HARRISON, M., NIENHUIS, A.N. (1982). Intronless human dihydrofolate reductase genes are derived from processed RNA molecules. P.N.A.S. 79, 7435-7439.

- CHENG, S.M. & SCHILDKRAUT, C.L. (1980). A family of moderately repetitive sequences in mouse DNA. Nucl.Acids Res. 8, 4075-4089.
- CHILDS, G., MAXSON, R., COHN, R.H., & KEDES, L. (1981). Orphans: Dispersed genetic elements derived from tandem repetitive genes of eukaryotes. Cell 23, 651-663.
- CHUNG, S., ZUKER, C. & LODISH, H.F. (1983). A repetitive and apparently transposable DNA sequence in Dictyostelium discoideum associated with developmentally regulated RNAs. Nucl. Acids Res. 11, 4835-4852.
- CLARKE, L. & CARBON, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature 287, 504-509.
- CLEWELL, D.B. & HELINSKI, D.R. (1969). Supercoiled circular DNA-protein complex in E.coli: Purification and induced conversion to an open circular DNA form. P.N.A.S. 62, 1159-1166.
- COEN, E., STRACHAN, T. & DOVER, G. (1982). Dynamics of concerted evolution of ribosomal DNA and Histone gene families in the melanogaster species subgroup of Drosophila. J.Mol.Biol. 158, 17-35.
- COLLINS, M. & RUBIN, G.M. (1982). Structure of the Drosophila Mutable Allele, white-crimson, and its white-ivory and Wild-type derivatives. Cell 30, 71-79.
- CRAIN, W.R., EDEN, F.C., PEARSON, W.R., DAVIDSON, E.H., & BRITTEN, R.J. (1976). Absence of short-period interspersal of repetitive and non-repetitive sequences in the DNA of Drosophila melanogaster. Chromosoma 56, 309-326.
- CREMONINI, R., LO-IERO, M. & DURANTE, M. (1981). Differentiation of Metaxylem cell line in the root of Allium cepa. II changes in nucleotide sequence redundancy. Protoplasma 108, 1-7.
- CULLIS, C.A. (1981). DNA sequence organisation in the flax genome. Biochim.Biophys.Acta, 652 1-15.

- CULLIS, C.A. & DAVIES, D.R. (1975). Ribosomal DNA amounts in Pisum sativum. Genetics 81, 485-492.
- CUNNINGHAM, M. (1983). Spot blot: A hybridisation assay for specific DNA sequences in multiple samples. Anal.Biochem. 128, 415-421.
- DAVIDSON, E.H. & BRITTEN, R.J. (1979). Regulation of Gene expression: Possible role of repetitive sequences. Science 204, 1052-1059.
- DAVIDSON, E.H., HOUGH, B.R., AMENSEN, C.S., & BRITTEN, R.J. (1973). General interspersion of repetitive and non-repetitive sequence elements in the DNA of Xenopus. J.Mol.Biol. 77, 1-23.
- DAVIDSON, E.H., KLEIN, W.H. & BRITTEN, R.J. (1977). Sequence organisation in animal DNA and a speculation on Hn RNA as a co-ordinate regulatory transcript. Dev.Biol. 55, 69-84.
- DAVIDSON, E.H., & POSAKONY, J.W. (1982). Repetitive sequence transcripts in development. Nature 297, 633-635.
- DEININGER, P.L., JOLLY, D.J., RUBIN, C.M., FRIEDMANN, T. & SCHMID, C.W. (1981). Base sequence studies of 300 nucleotide renatured repeated Human DNA clones. J.Mol.Biol. 151, 17-33.
- DENISON, R.A., VAN ARSDELL, S.W., BERNSTEIN, L.B. & WEINER, A.M. (1981). Abundant pseudo genes for small nuclear RNAs are dispersed in the human genome. P.N.A.S. 78, 810-814.
- DOOLITTLE, W.F. (1982). Selfish DNA after fourteen months. In: Genome Evolution (Dover, G.A. & Flavell, R.B., eds) pp. 3-28. Academic Press, London.
- DOOLITTLE, W.F. & SAPIENZA, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature 284, 601-603.
- DOVER, G. (1980). Ignorant DNA? Nature 285, 618-620.
- DOVER, G. (1982). Molecular drive: a cohesive mode of species evolution. Nature 299, 111-117.

- DOVER, G., BROWN, S., COEN, E., DALLAS, J., STRACHAN, T. & TRICK, M. (1982). The dynamics of genome evolution and species differentiation. In Genome Evolution (Dover, G.A. & Flavell, R.B., eds) pp. 343-372, Academic Press, London.
- DUBNICK, M., CHOU, J., PETES, T.D. & FARBER, R.A. (1983). Relationships among DNA sequences of the 1.3 kb Eco RI family of mouse DNA. J.Mol.Evol. 19, 115-121.
- DUNCAN, C., BIRO, P.A., CHOUDARY, P.U., ELDER, J.T., WANG, R.R.C., FORGET, B.C., DERIEL, J.K., & WEISSMAN, S.M. (1979). RNA polymerase III transcriptional units are interspersed among human non- $\alpha$ -globin genes. P.N.A.S. 76, 5095-5099.
- EDEN, F.C. & HENDRICK, J.P. (1978). Unusual organisation of DNA sequences in the chicken. Biochemistry 17, 5838-5844.
- EFSTRATIADIS, A., CRAIN, W.R., BRITTEN, R.J., DAVIDSON, E.H., & KAFATOS, F.C. (1976). DNA sequence organisation in the lepidopteran Antheraea pernyi. P.N.A.S., 73, 2289-2293.
- ELDER, J.T., PAN, J., DUNCAN, C.H. & WEISSMAN, S.M. (1981). Transcriptional analysis of interspersed repetitive polymerase III transcription units in human DNA. Nucl.Acids Res. 9, 1171-1189.
- EMMONS, S.W., YESNER, L., RUAN, K. & KATZENBERG, D. (1983). Evidence for a transposon in Caenorhabditis elegans. Cell 32, 55-65.
- EPPLEN, J.T., DIEDRICH, U., WAGENMANN, M., SCHMIDTKE, J. & ENGEL, W. (1979). Contrasting DNA sequence organisation patterns in Sauropsidian genomes. Chromosoma 75, 199-214.
- EPPLEN, J.T., LEIPOLDT, M., ENGEL, W. & SCHMIDTKE, J. (1978). DNA sequence organisation in Avian genomes. Chromosoma 69, 307-321.
- EVANS, I.J., JAMES, A.M. & BARNES, S.R. (1983). The organisation and evolution of repeated DNA sequences in closely-related plant genomes. J.Mol.Biol. 170 (in press).

- FINCHAM, J.R.S. (1983) "Genetics", J.Wright & Sons, Bristol.
- FINNEGAN, D.J., WILL, B.H., BAYEV, A.A., BOWCOCK, A.M., & BROWN, L. (1982). Transposable DNA sequences in eukaryotes. In: Genome Evolution (Dover, G.A. & Flavell, R.B., eds) pp.29-40. Academic Press, London.
- FIRTEL, R.A. & KINDLE, K. (1975). Structural organisation of the genome of the cellular slime-mold Dictyostelium discoideum - Inter-spersion of repetitive and single-copy DNA sequences. Cell 5, 401-411.
- FLAVELL, A.J. & ISH-HOROWICZ, D. (1981). Extra-chromosomal circular copies of the eukaryotic transposable element copia in cultured Drosophila cells. Nature 292, 591-595.
- FLAVELL, R.B. (1980). The molecular characterisation and organisation of plant chromosomal DNA sequences. Ann.Rev.Plant Physiol. 31, 569-596.
- FLAVELL, R.B. (1982). Sequence amplification, deletion and rearrangement: Major sources of variation during species divergence. In Genome Evolution (Dover, G.A. & Flavell, R.B. eds) pp.301-323. Academic Press, London.
- FLAVELL, R.B., BENNETT, M.D., SMITH, J.B. & SMITH, D.B. (1974). Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem.Genet. 12, 257-269.
- FLAVELL, R.B. O'DELL, M. & HUTCHINSON, J. (1981). Nucleotide sequence organisation in plant chromosomes and evidence for sequence translocation during evolution. C.S.H.S.Q.B. 45, 501-508.
- FLAVELL, R.B., RIMPAU, J. & SMITH, D.B. (1977). Repeated sequence DNA relationships in four cereal genomes. Chromosoma 63, 205-222.
- FLAVELL, R.B. & SMITH, D.B. (1974). Variation in nucleolar organiser rRNA gene multiplicity in wheat and rye. Chromosoma 47, 327-334.
- FLAVELL, R.B. & SMITH, D.B. (1976). Nucleotide sequence organisation in the wheat genome. Heredity 37, 231-252.
- FLAVELL, R.B. & SMITH, D.B. (1977). Hyperpolymer formation during renaturation of DNA from genomes with different sequence organisation. Nucl.AcidsRes. 4, 2429-2444.

- FRENCH, C.K. & MANNING, J.E. (1980). DNA sequence organisation in the Thysanura Thermobia domestica. J.Mol.Evol. 15, 277-289.
- FRIEDMAN, B.E., BOUCHARD, R.A. & STERN, H. (1982). DNA sequences repaired at pachytene exhibit strong homology among distantly related higher plants. Chromosoma 87, 409-424.
- FRITSCH, E.F., SHEN, C.K.J., LAWN, R.M. & MANIATIS, T. (1980). The organisation of repetitive sequences in mammalian globin gene clusters. C.S.H.S.Q.B. 45, 761-775.
- FRY, K. & SALSER, W. (1977). Nucleotide sequences of HS $\alpha$  satellite DNA from kangaroo rat Dipodomys ordii and characterisation of similar sequences in other rodents. Cell 12, 1069-1084.
- FUSSELL, C.P. (1975). The position of interphase chromosomes and late replicating DNA in centromere and telomere regions of Allium cepa L. Chromosoma 50, 201-210.
- GEHRING, W.J. & PARO, P. (1980). Isolation of a hybrid plasmid with homologous sequences to a transposing element of Drosophila melanogaster. Cell 19, 897-904.
- GEORGIEV, G.P. (1969). On the structural organisation of operon and the regulation of RNA synthesis in animal cells. J.Theor. Biol. 25, 473-490.
- GEORGIEV, G.P. ILYIN, Y.V., CHMELIAVSKAITE, V.G., RYSKOV, A.P., KRAMEROV, D.A., SKRYABIN, K.G., KRAYEV, A.S., LUKANIDIN, E.M., GRIGORYAN, M.S. (1981). Mobile dispersed genetic elements and other middle repetitive DNA sequences in the genomes of Drosophila and mouse: Transcription and Biological significance. C.S.H.S.Q.B. 45, 641-654.
- GOLDBERG, R.B., HOSCHEK, G. & VODKIN, L.O. (1983). An insertion sequence blocks the expression of a soybean lectin gene. Cell 33, 465-475.

- GOLDSCHMIDT, R. (1940). "The material basis of evolution." Yale University Press.
- GOULD, S.J. & ELDRIDGE, N. (1977). Punctuated equilibria : the tempo and mode of evolution reconsidered. Palaeobiology 3, 23-40.
- GRAHAM, D.E., NEUFELD, B.R., DAVIDSON, E.H. & BRITTEN, R.J. (1974). Interspersion of repetitive and non-repetitive DNA sequences in the sea urchin genome. Cell 1, 127-137.
- GRIMALDI, G., QUEEN, C., & SINGER, M.F., (1981). Interspersed repeated sequences in the African Green Monkey that are homologous to the human Alu family. Nucl.Acids Res. 9, 5553-5568.
- GRIMALDI, G. & SINGER, M.F. (1982). A monkey Alu sequence is flanked by 13-base-pair direct repeats of an interrupted  $\alpha$  - satellite DNA sequence. P.N.A.S. 79, 1497-1500.
- GRIME, J.P. & MOWFORTH, M.A. (1982). Variation in genome size - an ecological interpretation. Nature 299, 151-153.
- GRUNSTEIN, M. & HOGNESS, D.S. (1975). Colony hybridisation: A method for the isolation of cloned DNA's that contain a specific gene. P.N.A.S. 72, 3961-3965.
- GUPTA, V.S., GADRE, S.R. & RANJEKAR, P.K. (1982) Novel DNA sequence organisation in rice genome. Biochim.Biophys. Acta 656, 147-154.
- GURLEY, W.B., HEPBURN, A.G. & KEY, J.L. (1979). Sequence organisation of the soybean genome. Biochim.Biophys Acta. 561, 167-183.
- HAIGWOOD, N.L., JAHN, C.L., HUTCHINSON, C.A. & EDGEELL, M.H. (1981). Locations of three repetitive sequence families found in BALB/c adult  $\beta$ -globin clones. Nucl.Acids Res. 9: 1133-1150.
- HAKE, S. & WALBOT, V. (1980). The genome of Zea mays: its organisation and homology to related grasses. Chromosoma 79, 251-270.



- HATCH, F.T., BODNER, A.J., MAZRIMAS, J.A., & MOORE, D.H. (1976). Satellite DNA and cytogenetic evolution: DNA quantity, satellite DNA and karyotypic variation in kangaroo rats (Genus Dipodomys). Chromosoma 58, 155-168.
- HELLER, R. & ARNHEIM, N. (1980). Structure and organisation of the highly repeated and interspersed 1.3 kb Eco R.I. - Bgl II sequence family in mice. Nucl.Acids Res. 8, 5031-5043.
- HESLOP-HARRISON, J.S. & BENNETT, M.D. (1983a). Prediction and analysis of spatial order in haploid chromosome complements. Proc. R.Soc. Lond. 218, 211-223.
- HESLOP-HARRISON, J.S. & BENNETT, M.D. (1983b). The spatial order of chromosomes in root-tip metaphases of Aegilops umbellulata. Proc.R.Soc.Lond. 218, 225-239.
- HORZ, W. & ZACHAU, H.G. (1977). Characterisation of distinct segments in mouse satellite DNA by restriction nucleases. Eur.J. Biochem. 73, 383-392.
- HOUCK, C.M., RINEHART, F.P., & SCHMID, C.W. (1979). A ubiquitous family of repeated DNA sequences in the human genome. J.Mol.Biol. 132, 289-306.
- HOUCK, C.M. & SCHMID, C.W. (1981). The evolution of a family of short interspersed repeats in primate DNA. J.Mol.Evol. 17, 148-155.
- HOURCADE, D., DRESSLER, D. & WOLFSON, J. (1973). The amplification of ribosomal RNA genes involves a rolling circle intermediate. P.N.A.S. 70, 2926-2930.
- HUDSPETH, M.E.S., TIMBERLAKE, W.E. & GOLDBERG, R.B. (1977). DNA sequence organisation in the water mold Achlya. P.N.A.S. 74, 4332-4336.
- ILYIN, Y.V., CHMELIAUSKAITE, V.G., GEORGIEV, G. (1980). Double stranded sequences in RNA of Drosophila melanogaster: relation to mobile dispersed genes. Nucl.Acids Res. 8, 3439-3457.

- INGLE, J., PEARSON, G.G. & SINCLAIR, J. (1973). Species distribution and properties of nuclear satellite DNA in higher plants. Nature New Biol. 242, 193-197.
- INGLE, J. TIMMIS, J.N. & SINCLAIR J. (1975). The relationship between satellite DNA, rRNA gene redundancy and genome size in plants. Plant physiol. 55, 496-501.
- ISH-HOROWICZ, D. & BURKE, J.F. (1981). Rapid and efficient cosmid cloning. Nucl.Acids Res. 9, 2989-2999.
- JACKSON, J.A. & FINK, G.R. (1981). Gene conversion between duplicated genetic elements in yeast. Nature 292, 306-311.
- JAGADEESWARAN, P., FORGET, B.G. & WEISSMAN, S.M. (1981). Short interspersed repetitive DNA elements in Eukaryotes: Transposable DNA elements generated by reverse transcription of RNA pol III transcripts Cell 26, 141-142.
- JAGADEESWARAN, P., TUAN, D., FORGET, B.G. & WEISSMAN, S.M. (1982). A gene deletion ending at the midpoint of a repetitive DNA sequence in one form of hereditary persistence of fetal haemoglobin. Nature 296, 469-470.
- JEFFREYS, A.J. (1982). Evolution of globin genes. In: Genome Evolution (Dover, G.A. & Flavell, R.B., eds) pp.157-176. Academic Press, London.
- JELINEK, W.R. & SCHMID, C.W. (1982). Repetitive sequences in eukaryotic DNA and their expression. Ann.Rev.Biochem. 51, 813-844.
- JELINEK, W.R., TOOMEY, T.P., LEINWAND, L., DUNCAN, C.H., BIRO, P.A., CHOUDARY, P.V., WEISSMAN, S.M., RUBIN, C.M., HOUCK, C.M., DEININGER, P.L., & SCHMID, C.W. (1980). Ubiquitous, interspersed repeated sequences in mammalian genomes. P.N.A.S. 77, 1398-1402.
- JOHN, B. & MIKLOS, G.L.G. (1979). Functional aspects of satellite DNA and heterochromatin. Int.Rev.Cytol. 58, 1-114.

- JONES, H.A. & MANN, L.K. (1963).  
In: "Onions and their allies",  
(Jones, H.A. & Mann, L.K. eds).  
Interscience, New York.
- JONES, J.D.G. & FLAVELL, R.B. (1982). The  
structure amount and chromosomal localisa-  
tion of defined repeated DNA sequences in  
species of the genus Secale.  
Chromosoma 86, 613-641.
- JONES, R.N. & REES, H. (1969). Nuclear DNA  
variation in Allium. Heredity 23,  
591-605.
- JUDD, B.H., SHEN, M.W. & KAUFMAN, T.C. (1972).  
The anatomy and function of a segment  
of the x-chromosome of D.melanogaster.  
Genetics 71, 139-156.
- KIPER, M. & HERZFELD, F. (1978). DNA sequence  
organisation in the genome of  
Petroselinum sativum (Umbelliferae).  
Chromosoma 65, 335-351.
- KIRK, J.T.O., REES, H. & EVANS, G. (1970).  
Base composition of nuclear DNA  
within the genus Allium. Heredity 25,  
507-512.
- KONVICKA, O. & LEVAN, A. (1972). Chromosome  
studies in Allium sativum. Hereditas  
72, 129-148.
- KRAMEROV, D. A. LEKAKH, I.V., SAMARINA, O.P.,  
& RYSKOV, A.P. (1982). The sequences  
homologous to major interspersed  
repeats B1 and B2 of mouse genome  
are present in mRNA and small cyto-  
plasmic poly (A)<sup>+</sup> RNA.  
Nucl.Acids Res. 10, 7477-7491.
- KRAYEV, A.S. KRAMEROV, D.A., SKRYABIN, K.G.,  
RYSKOV, A.P., BAYEV, A.A. & GEORGIEV, G.P.  
(1980). The nucleotide sequence of the  
ubiquitous repetitive DNA sequence B1  
complementary to the most abundant class  
of mouse fold-back RNA. Nucl. Acids  
Res. 8, 1201-1215.
- KROLEWSKI, J.J., BERTELSEN, A.H., HUMAYUN, M.Z. &  
RUSH, M.G. (1982). Members of the  
Alu family of interspersed repetitive  
DNA sequences are in the small circular  
DNA population of monkey cells grown in  
culture. J.Mol.Biol. 154, 399-415.

- KRUMLAUF, R. & MARZLUF, G.A. (1979). Characterisation of the sequence complexity and organisation of the Neurospora crassa genome. Biochemistry **18**, 3705-3713.
- LASKEY, R.A. (1980). The use of intensifying screens or organic scintillators for visualising radioactive molecules resolved by gel electrophoresis. In: Methods in Enzymology **65**. (Grossman, L. & Moldave K., eds), pp.363-371 Academic Press, New York.
- LERMAN, M.I., THAYER, R.E. & SINGER M.F. (1983). Kpn I family of long interspersed repeated DNA sequences in primates: Polymorphism of family members and evidence for transcription. P.N.A.S. **80**, 3966-3970.
- LERNER, M.R., BOYLE, J.A., MOUNT, S.M., WOLIN, S.L. & STEITZ, J.A. (1980). Are snRNPs involved in splicing? Nature **283**, 220-224.
- LERNER, M.R. & STEITZ, J.A. (1981). Snurps and Scryps. Cell **25**, 298-300.
- LEVAN, A. (1935). Cytological studies in Allium:VI, The chromosome morphology of some diploid species of Allium. Hereditas **20**, 289-330.
- LEVIS, R., DUNSMUIR, P. & RUBIN, G.M. (1980). Terminal repeats of the Drosophila transposable element copia: nucleotide sequence and genomic organisation. Cell **21**, 581-588.
- LEWIN, B. (1980). "Gene Expression 2". (Lewin, B. ed) 2nd Edition. J.Wiley & Sons Ltd., New York.
- LEWIN, R. (1982). Repeated DNA still in search of a function. Science **217**, 621-623.
- LIU, L.S. & LARK, K.G. (1982). The Red function of phage $\lambda$  mediates the alteration of an interspersed repeated DNA sequence from the kangaroo rat Dipodomys ordii. Mol.Gen.Genet. **188**, 27-36.
- MACGREGOR, H.C. (1982). Chromosomes and speciation in Amphibia. In: Genome Evolution (Dover, G.A. & Flavell, R.B., eds) pp.325-341 Academic Press, London.
- MAGGINI, F., BARSANTI, P. & MARAZIA, T. (1978). Individual variation of the nucleolus organizer regions in Allium cepa and A. sativum. Chromosoma **66**, 173-183.

- MANIATIS, T., FRITSCH, E.F. & SAMBROOK, J. (1982). Molecular cloning-a laboratory manual Cold Spring Harbor Laboratory.
- MANNING, J.E., SCHMID, C.W. & DAVIDSON, N. (1975). Interspersion of repetitive and non-repetitive DNA sequences in the Drosophila melanogaster genome. Cell 4, 141-155.
- MANUELIDIS, L. (1982). Repeated DNA sequences and nuclear structure. In Genome Evolution (Dover, G.A. & Flavell, R.B. eds) pp.263-285. Academic Press, London.
- MANUELIDIS, L. & BIRO, P.A. Genomic representation of the Hind III 1.9 kb repeated DNA. Nucl. Acids Res. 10, 3221-3239.
- MARMUR, J. & DOTY, P. (1959). Heterogeneity in DNA.1:Dependence on composition of the configurational stability of DNAs. Nature 183, 1427-1428.
- MARX, J.L. (1983). A transposable element of Maize emerges. Science 219, 829-830.
- MAYNARD-SMITH, J. (1982). Overview - Unsolved Evolutionary problems. In: Genome Evolution (Dover, G.A. & Flavell, R.B. eds) pp. 375-382. Academic Press, London.
- MCCLINTOCK, B. (1956). Controlling elements and the gene. C.S.H.S.Q.B. 21, 197-216.
- MCGINNIS, W. & BECKENDORF, S.K. (1983). Association of a Drosophila transposable-element of the roo family with chromosomal deletion breakpoints. Nucl.Acids Res. 11 737-751.
- MEUNIER-ROTIVAL, M., SORIANO, P., CUNY, G., STRAUSS, F., & BERNADI, G. (1982). Sequence organisation and genomic distribution of the major family of interspersed repeats of mouse DNA. P.N.A.S. 79, 355-359.
- MIKLOS, G.L.G. (1982). Sequencing and manipulating highly repeated DNA. In: Genome Evolution (Dover, G.A. & Flavell, R.B., eds). pp.41-68. Academic Press, London.

- MIKLOS, G.L.G. & GILL, A.C. (1981). The DNA sequences of cloned complex satellite DNAs from Hawaiian Drosophila and their bearing on satellite DNA sequence conservation. Chromosoma 82, 409-427.
- MIKLOS, G.L.G., WILLCOCKS, D.A. & BAVERSTOCK, P.R. (1980). Restriction endonuclease and molecular analyses of three rat genomes with special reference to chromosome rearrangement and speciation problems. Chromosoma 76, 339-363.
- MODOLELL, J., BENDER, W. & MESELSON, M. (1983). Drosophila melanogaster mutations suppressible by the suppressor of Hairy-wing are insertions of a 7.3 kb mobile element. P.N.A.S. 80, 1678-1682.
- MOORE, G.P., PEARSON, W.R., DAVIDSON, E.H. & BRITTEN, R.J. (1981). Long and short repeats of sea urchin DNA and their evolution. Chromosoma 84, 19-32.
- MOYZIS, R.K., BONNET, J., LI, D.W. & TS'0, P.O.P. (1981). An alternative view of mammalian DNA sequence organisation. I Repetitive sequence interspersion in Syrian Hamster DNA: A model system. J.Mol.Biol. 153, 841-870.
- MURRAY, M.G., CUELLAR, R.E. & THOMPSON, W.F. (1978) DNA sequence organisation in the pea genome. Biochemistry 17, 5781-5790.
- MURRAY, M.G., PALMER, J.D., CUELLAR, R.E. & THOMPSON, W.F. (1979). Deoxyribonucleic acid sequence organisation in the Mung bean genome. Biochemistry 18, 5259-5266.
- MURRAY, V. & HOLLIDAY, R. (1979). A mechanism for RNA-RNA splicing and a model for the control of gene expression. Genet. Res. (Camb). 34, 173-188.
- MUSICH, P.R., MAIO, J.J. & BROWN, F.L. (1977). Subunit structure of chromatin and the organisation of eukaryotic highly repetitive DNA: indications of a phase relation between restriction sites and chromatin subunits in African green monkey and calf nuclei. J.Mol.Biol. 117, 657-678.

- MUSTI, A.M., SOBIESKI, D.A., CHEN, B.B. & EDEN, F.C., (1981). Repeated Deoxyribonucleic Acid clusters in the chicken genome contain homologous sequence elements in scrambled order. Biochemistry 20, 2989-2999.
- NAGL, W. (1982). DNA Endoreduplication and differential replication. In: Encyclopaedia of Plant Physiology 14B. (Boulter, D. & Parthier, B., eds). pp. 111-124. Springer-Verlag, Berlin.
- NAGYLAKI, T. & PETES, T.D. (1982). Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics 100, 315-337.
- NARAYAN, R.K.J. (1982). Discontinuous DNA variation in the evolution of plant species: The genus Lathyrus. Evolution 36, 877-891.
- NARAYAN, R.K.J. & DURRANT, A. (1983). DNA distribution in chromosomes of Lathyrus species. Genetica 61, 47-53.
- NARAYAN, R.K.J. & REES, H. (1976). Nuclear DNA variation in Lathyrus. Chromosoma 54, 141-154.
- NARAYAN, R.K.J. & REES, H. (1977). Nuclear DNA divergence among Lathyrus species. Chromosoma 63, 101-107.
- OHNO, S. (1971). Simplicity of mammalian regulatory systems inferred by single gene determination of sex phenotypes. Nature 234, 134-137.
- OHTA, T. (1983). Theoretical study on the accumulation of selfish DNA. Genet.Res. (Camb) 41, 1-15.
- ORGE, L.E. & CRICK, F.H.C. (1980). Selfish DNA: the ultimate parasite. Nature 284, 604-607.
- PARO, R., GOLDBERG, M.L. & GEHRING, W.J. (1983). Molecular analysis of large transposable elements carrying the white locus of Drosophila melanogaster. EMBO Journal 2, 853-860.

- PERLMAN, J. (1983). One side of a deletion breakpoint from the Drosophila melanogaster genome contains a transposable element. Gene 21, 87-94.
- POSAKONY, J.W., FLYTZANIS, C.N., BRITTEN, R.J. & DAVIDSON, E.H. (1983). Interspersed sequence organisation and developmental representation of cloned poly (A) RNAs from Sea Urchin Eggs. J.Mol.Biol. 167, 361-389.
- POSAKONY, J.W., SCHELLER, R.H., ANDERSON, D.M., BRITTEN, R.J. & DAVIDSON, E.H. (1981). Repetitive sequences of the sea urchin genome. III Nucleotide sequences of cloned repeat elements. J.Mol.Biol. 149, 41-67.
- RANJEKAR, P.K., PALLOTTA, D. & LAFONTAINE, J.G. (1978). Analysis of plant genomes. V Comparative study of molecular properties of DNAs of seven Allium species. Biochem.genet. 16, 957-971.
- REES, H. & JONES, R.N. (1977). Chromosome Genetics. Edward Arnold Ltd., London.
- RIGBY, P.W.J., DIECKMANN, M., RHODES, C. & BERG, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by Nick translation with DNA polymerase I. J.Mol.Biol. 113, 237-251.
- RIMPAU, J., SMITH, D. & FLAVELL, R. (1978). Sequence organisation analysis of the Wheat and Rye genomes by Interspecies DNA/DNA hybridisation. J.Mol.Biol. 123, 327-359.
- RIMPAU, J., SMITH, D. & FLAVELL, R. (1980). Sequence organisation in barley and oats chromosomes revealed by interspecies DNA/DNA hybridisation. Heredity 44, 131-149.
- ROBERTS, J.M., BUCK, L.B. & AXEL, R. (1983). A structure for amplified DNA. Cell 33, 53-63.
- ROBERTS, R. (1980). Small RNA's and splicing. Nature 283, 132-133.



- ROEDER, G.S., FARABAUGH, P., CHALEFF, D. & FINK, G. (1980). The origins of gene instability in yeast. Science **209**, 1375-1380.
- ROIZES, G.P. & PAGES, M. (1982). Conserved and divergent sequences of Bovine satellite DNA's. In: Genome Evolution (Dover, G.A. & Flavell, R.B., eds). pp. 95-111. Academic Press, London.
- ROSE, M.R. & DOOLITTLE, W.F. (1983). Molecular biological mechanisms of speciation. Science **220**, 157-162.
- RUBIN, G.M., BROREIN, W.J., DUNSMUIR, P., FLAVELL, A.J., LEVIS, R., STROBEL, E., TOOLE, J.J. & YOUNG, E. (1981). Copia-like transposable elements in the Drosophila genome. C.S.H.S.Q.B. **45**, 619-628.
- RUBIN, G.M., KIDWELL, M.G. & BINGHAM, P.M. (1982). The Molecular Basis of P-M Hybrid Dysgenesis: The Nature of Induced Mutations. Cell **29**, 987-994.
- RYFFEL, G.U., MUELLENDER, D.B., WYLER, T., WAHLI, W. & WEBER, R. (1981). Transcription of single-copy vitellogenin gene of Xenopus involves expression of middle repetitive DNA. Nature **291**, 429-431.
- SCHELLER, R.H., MCALLISTER, L.B., CRAIN, W.R., DURICA, D.S., POSAKONY, J.W., THOMAS, T.L. BRITTEN, R.J. & DAVIDSON, E.H. (1981). Organisation and expression of multiple actin genes in the sea urchin. Mol. Cell. Biol. **1**, 609-628.
- SCHERER, S. & DAVIS, R.W. (1980). Recombination of dispersed repeated DNA sequences in yeast. Science **209**, 1380-1384.
- SCHIMKE, R.T. (1982). "Gene Amplification" C.S.H.L., New York.
- SCHMID, C.W. & DEININGER, P.L. (1975). Sequence organisation of human genome. Cell **6**, 345-358.
- SCHMIDTKE, J., EPPLEN, J.T. & ENGEL, W. (1979). Genome analysis of amphioxus and speculation as to the origin of contrasting vertebrate genome organisation patterns. Comp. Biochem. Physiol. **63B**, 455-458.

- SEAL, A.G. & REES, H. (1982). The distribution of quantitative DNA changes associated with the evolution of diploid Festuceae. Heredity 49, 179-190.
- SHAFIT-ZAGARDO, B., BROWN, F.L., ZAVODNY, P.J. & MAIO, J.J. (1983). Transcription of the Kpn 1 families of long interspersed DNAs in human cells. Nature 304, 277-280.
- SHARP, P.A. (1983). Conversion of RNA to DNA in mammals: Alu-like elements and pseudogenes. Nature 301, 471-472.
- SHIBA, T. & SAIGO, K. (1983). Retrovirus-like particles containing RNA homologous to the transposable element copia in Drosophila melanogaster. Nature 302, 119-124.
- SINGER, M.F. (1982). Highly repeated sequences in mammalian genomes. Int.Rev.Cytol. 76, 67-112.
- SINGER, M.F., THAYER, R.E., GRIMALDI, G., LERMAN, M.I. & FANNING, T.G. (1983). Homology Between the kpn I primate and Bam HI (MIF-1) rodent families of long interspersed repeated sequences. Nucl.Acids Res. 11, 5739-5745.
- SMITH, D.B. & FLAVELL, R.B. (1977). Nucleotide sequence organisation in the rye genome. Biochim.Biophys Acta 474, 82-97.
- SMITH, D.B., RIMPAU, J. & FLAVELL, R.B. (1976). Interspersion of different repeated sequences in the wheat genome revealed by interspecies DNA/DNA hybridisation. Nucl.Acids Res. 3, 2811-2825.
- SMITH, G.P. (1976). Evolution of repeated DNA sequences by unequal crossovers. Science 191, 528-535.
- SNYDER, M.P., KIMBRELL, D., HUNKAPILLER, M., HILL, R., FRISTROM, J. & DAVIDSON, N. (1982). A transposable element that splits the promoter region inactivates a Drosophila cuticle protein gene. P.N.A.S. 79, 7430-7434.

- SOUTHERN, E.M. (1975a). Long range periodicities in mouse satellite DNA. J.Mol.Biol. 94, 51-70.
- SOUTHERN, E.M. (1975b). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J.Mol.Biol. 98, 503-518.
- SPARROW, A.H. & NAUMAN, A.F. (1976). Evolution of genome size by DNA doublings. Science 192, 524-529.
- STACK, S.M. & COMINGS, D.E. (1979). The chromosomes and DNA of Allium cepa. Chromosoma 70, 161-181.
- STANLEY, S.M. (1979). "Macroevolution." Freeman, San Francisco.
- STEARNS, W.T. (1944). Notes on the genus Allium in the old world. Herbertia 11, 11-34.
- STEBBINS, G.L. (1982). Plant Speciation. In: Mechanisms of speciation (Barigozzi, ed) pp.21-40. Liss, A.R., New York.
- STRACHAN, T., COEN, E., WEBB, D. & DOVER G. (1982). Modes and rates of change of complex DNA families of Drosophila. J.Mol.Biol. 158, 37-54.
- STRATHERN, J.N., KLAR, A.J.S., HICKS, J.B., ABRAHAM, J.A., IVY, J.M., NASMYTH, K.A. & MCGILL, C. (1982). Homothallic switching of yeast Mating-Type Cassettes is initiated by a double-stranded Cut in the MAT locus. Cell 31, 183-192.
- STREECK, R.E. (1982). A multicopy insertion sequence in the bovine genome with structural homology to the long terminal repeats of retroviruses. Nature 298, 767-769.
- STROBEL, E., DUNSMUIR, P. & RUBIN, G.M. (1979). Polymorphisms in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. Cell 17, 429-439.
- SUEOKA, N. & CHENG, T.Y. (1962). Natural occurrence of a DNA resembling the deoxyadenylate-deoxythymidylate polymer. P.N.A.S. 48, 1851-1856.

- TARTOF, K.D. (1975). Redundant Genes. Ann.Rev.Genet. 9, 355-385.
- TASHIMA, M., CALABRETTA, B., TORELLI, G., SCOFIELD, M., MAIZEL, A., & SAUNDERS, G.F. (1981). Presence of a highly repetitive and widely dispersed DNA sequence in the human genome. P.N.A.S. 78, 1508-1512.
- THOMPSON, W.F. & MURRAY, M.G. (1981). The nuclear genome: Structure and function. In: The Biochemistry of Plants. Vol.6 (Marcus, A. ed) pp. 1-81. Academic Press, New York.
- TIMBERLAKE, W.E. (1978). Low repetitive DNA content in Aspergillus nidulans. Science 202, 973-975.
- TRAUB, H.P. (1968). The subgenera, sections and subsections of Allium L. Plant Life 24, 147-163.
- TRUETT, M.A., JONES, R.S. & POTTER, S.S. (1981). Unusual structure of the FB family of transposable elements in Drosophila. Cell 24, 753-763.
- TUTIN, T.G. (1980). Flora Europaea, Vol.5). Cambridge Univ.Press, Cambridge.
- VARLEY, J.M., MACGREGOR, H.C. & ERBA, H.P. (1980). Satellite DNA is transcribed on lampbrush chromosomes. Nature 283, 686-688.
- VED BRAT, S. (1965a). Genetic systems in Allium I Chromosome variation. Chromosoma 16, 486-499.
- VED BRAT, S. (1965b). Genetic systems in Allium III Meiosis and breeding systems. Heredity 20, 325-339.
- VERMA, S.C. & MITTAL, R.K. (1978). Chromosome variation in the common garlic. Allium sativum L. Cytologia 43, 383-396.
- VOSA, C.G. (1966). Seed germination and B-chromosomes in the Leek (Allium porrum). In: Chromosomes Today, Vol. 1, (Darlington, C.D. & Lewis, K.R. eds). pp.24-30. Oliver & Boyd, Edinburgh.

- WAHL, G.M., STERN, M. & STARK, G.R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridisation by using dextran sulphate. P.N.A.S. 76, 3683-3687.
- WALBOT, V. & DURE, L.S. (1976). Developmental biochemistry of cotton seed embryogenesis and germination. VII Characterisation of the cotton genome. J.Mol.Biol. 101, 503-536.
- WALBOT, V. & GOLDBERG, R. (1979). Plant genome organisation and its relationship to classical plant genetics. In: Nucleic acids in Plants. Vol I (Hall, T.C. & Davies, J.W., eds). pp.1-40. CRC Press.
- WALKER, P.M.B. (1968). How different are the DNAs from related animals? Nature 219, 228-232.
- WALKER, P.M.B. (1971). Origin of satellite DNA. Nature 229, 306-308.
- WELLS, R., ROYER, H.D. & HOLLENBERG, C.P. (1976). Non Xenopus-like DNA sequence organisation in the Chironomus tentans genome. Mol.Gen.Genet. 147, 45-51.
- WENSINK, P.C., TABATA, S. & PACHL, C. (1979). The clustered and scrambled arrangement of moderately repetitive elements in Drosophila DNA. Cell 18, 1231-1246.
- WHITE, M.J.D. (1982). Rectangularity, speciation and chromosome architecture. In: Mechanisms of speciation (Barigozzi, ed) pp.75-104. A.R. Liss, New York.
- WHITNEY, J.B. & LAMOREUX, M.L. (1982). Transposable elements controlling genetic instabilities in mammals. J.Heredit 73, 12-18.
- WILLIAMSON, V.M., COX, D. YOUNG, E.T., RUSSELL, D.W., & SMITH, M. (1983). Characterisation of transposable element-associated mutations that alter yeast alcohol dehydrogenase II expression. Mol.Cell.Biol. 3, 20-31.
- WIMPEE, C.F. & RAWSON, J.R.Y. (1979). Characterisation of the nuclear genome of pearl millet. Biochim.Biophys. Acta 562, 192-206.

- YAMAMOTO, M. (1979). Cytological studies of heterochromatin function in the Drosophila melanogaster male: Autosomal meiotic pairing. Chromosoma 72, 293-328.
- ZIEVE, G. & PENMAN, S. (1981). Subnuclear particles containing a small nuclear RNA and heterogeneous nuclear RNA. J.Mol.Biol. 145, 501-523.
- ZIMMERMAN, J.L. & GOLDBERG, R.B. (1977). DNA sequence organisation in the genome of Nicotiana tabacum. Chromosoma 59, 227-252.
- ZUKER, C. & LODISH, H.F. (1981). Repetitive DNA sequences cotranscribed with developmentally regulated Dictyostelium discoideum mRNAs. P.N.A.S. 78, 5386-5390.

PUBLICATIONS

Evans, I.J., James, A.M., & Barnes, S.R.

The organisation and evolution of repeated DNA  
sequences in closely-related plant genomes.

J.Mol.Biol. In Press

(Accepted August 1983).